

THÈSE PRÉSENTÉE
POUR OBTENIR LE GRADE DE
DOCTEUR DE
L'UNIVERSITÉ DE BORDEAUX

ÉCOLE DOCTORALE SP2 : Sociétés, Politique, Santé Publique

THESE EN SCIENCE POLITIQUE
SCIENCES PO BORDEAUX

Laboratoire d'accueil : Centre Emile Durkheim

Par Agathe DEVAUX-SPATARAKIS

**LA METHODE EXPERIMENTALE PAR ASSIGNATION
ALEATOIRE : UN INSTRUMENT DE RECOMPOSITION
DE L'INTERACTION ENTRE SCIENCES SOCIALES ET
ACTION PUBLIQUE EN FRANCE ?**

Sous la direction de : Thierry BERTHET

Soutenue le 6 octobre 2014

Membres du jury :

M. Andy SMITH, Directeur de recherche FNSP, Centre Emile Durkheim, Sciences Po Bordeaux ; Président

M. Claudio RADAELLI, Professeur, Université d'Exeter ; Rapporteur

M. Frédéric VARONE, Professeur, Université de Genève ; Rapporteur

M. Eric MONNIER, Professeur associé à l'Université de Genève ; Examineur

Mme. Agnès LABROUSSE, Maître de conférences, Université de Picardie ; Examinatrice

M. Thierry BERTHET, Directeur de recherche CNRS, Centre Emile Durkheim, Sciences Po Bordeaux; Directeur de thèse

Titre : La méthode expérimentale par assignation aléatoire : un instrument de recomposition de l'interaction entre sciences sociales et action publique en France ?

Résumé : Depuis le début des années 2000, le recours à la méthode expérimentale par assignation aléatoire pour évaluer les dispositifs publics connaît un essor mondial sans précédent. Cette méthode scientifique est présentée par ses promoteurs comme la plus rigoureuse pour estimer l'impact d'une intervention ainsi que la mieux à même de favoriser la prise en compte des preuves scientifiques par les décideurs politiques. Son utilisation dans le cadre de *l'evidence-based policy* nous amène à considérer cette méthode comme un instrument, une institution sociale, visant à organiser un apprentissage commun entre les acteurs de l'action publique et les acteurs scientifiques. L'observation de cette interaction constitue le cœur de ce travail de recherche. Ce dernier étudie comment l'inscription de ses parties prenantes dans leurs *champs d'action stratégiques* respectifs conditionne l'usage de cette méthode sur le territoire français. Celle-ci se décline alors, en une variété de *sites institutionnels* témoignant des oppositions entre les pratiques, les intérêts, et les modèles d'apprentissages des acteurs la composant. Notre analyse de 15 études de cas de son utilisation sur le territoire français, révèle les tensions inhérentes à l'application de cette méthode sur de nouveaux dispositifs d'intervention sociale, et questionne ses capacités à produire un apprentissage commun entre acteurs de l'action publique et acteurs scientifiques.

Mots clés : Evaluation des politiques publiques, evidence-based policy, expérimentations sociales, instruments d'action publique, apprentissage politique

Title : Randomized controlled trials: rearranging the interaction of social sciences and public action in France?

Abstract : The start of the XXIst century witnessed an unprecedented use of randomized controlled trials to assess public programs across the world. This scientific method has been championed as the most rigorous to assess the impact of public intervention. It has promoted the use of scientific evidence by policy makers through the evidence-based policy movement. Therefore, this method is to be understood as a social institution whose aim is to organize a joint learning between policy actors and scientists. The study of this interaction is the core of this dissertation, which analyzes how stakeholders' memberships to their respective strategic action fields drive this method's use in France. Through the execution of this method, we observe a range of different institutional sites. They are testimonies of the opposing practices, interests and learning patterns of the actors involved. This dissertation analyses 15 case studies of this method's implementation in France. It reveals the inherent tensions at work in its use on new social programs, and challenges this method's ability to produce a joint learning between policy actors and scientists.

Keywords : Evaluation of public policies, evidence-based policy, experiments, instrument of public action, political learning

Unité de recherche

Centre Emile Durkheim – science politique et sociologie comparatives,
UMR5116, Sciences Po Bordeaux, 11 allée Ausone 33607 Pessac Cedex

Sommaire

REMERCIEMENTS.....	9
LEXIQUE DES ACRONYMES.....	10
INTRODUCTION.....	11
1. L'EXPERIMENTATION ALEATOIRE, INSTRUMENT DE RECONCILIATION DES LOGIQUES SCIENTIFIQUES ET POLITIQUES ?	13
2. UN CADRE NEO-INSTITUTIONNALISTE SOCIOLOGIQUE POUR APPREHENDER UNE METHODE SCIENTIFIQUE ET SES EFFETS.....	16
3. UNE STRATEGIE DE RECHERCHE ITERATIVE ENTRE COLLECTE DE DONNEES ET THEORIE	20
3.1. <i>Le contrat Cifre, de la participation observante à l'observation participante.....</i>	20
3.2. <i>Appareillage de collecte de données</i>	23
3.3. <i>Une articulation du travail théorique et empirique par abduction.....</i>	24
3.4. <i>Un positionnement épistémologique de réalisme critique.....</i>	26
1^{ERE} PARTIE : UNE STRATEGIE DE RECHERCHE ELABOREE PAR ABDUCTION: UN CHOIX THEORIQUE ARTICULE A LA SOCIOGENESE D'UN OBJET D'ETUDE	28
1. <i>L'EVIDENCE-BASED POLICY</i> ET LA METHODE EXPERIMENTALE PAR ASSIGNATION ALEATOIRE: SOCIOGENESE DE DEUX OBJETS EN INTERACTION	31
1.1. <i>La méthode expérimentale par assignation aléatoire, un outil structurant de l'interaction entre la science et l'action publique</i>	34
1.1.1. Les origines d'un outil de production et de communication scientifique.....	35
1.1.2. Le développement de l'utilisation de la méthode expérimentale aléatoire pour l'aide à la prise de décision, l'émergence de « la méthode la plus rigoureuse »	44
1.2. <i>La méthode expérimentale par assignation aléatoire comme instrument de l'Evidence-Based Policy</i>	55

1.2.1.	Les incarnations de <i>l'Evidence-Based Policy</i> , les tensions entre modèle standardisé et contextes nationaux.....	56
1.2.2.	L'expérimentation aléatoire en débat : analyse des théories portées par l'expérimentation aléatoire comme instrument de l'EBP	71
2.	ETUDE D'UN INSTRUMENT D'ACTION PUBLIQUE PAR UNE APPROCHE CONSTRUCTIVISTE DE L'INSTITUTION, POSITIONNEMENT THEORIQUE ET IMPLICATIONS METHODOLOGIQUES	82
2.1.	<i>L'Etude empirique de l'émergence d'une méthode d'évaluation de politiques publiques par la théorie générale des champs et le travail institutionnel</i>	82
2.1.1.	Un cadrage théorique alliant la sociologie des champs et le néo-institutionnalisme sociologique.....	83
2.1.2.	Démarche empirique du chapitre 2.....	99
2.2.	<i>Appréhender la méthode randomisée dans sa diversité, une stratégie de recherche néo-institutionnaliste centrée sur l'apprentissage</i>	102
2.2.1.	Étudier la méthode randomisée comme un site institutionnel d'une relation clinique organisant un type particulier d'apprentissage	103
2.2.2.	Démarche empirique du chapitre 3.....	117
2.2.3.	Protocole de collecte de données.....	122
2^{EME}	PARTIE : L'INSTITUTIONNALISATION DE LA «METHODE EXPERIMENTALE » EN FRANCE: UN SUCCES TEMPORAIRE ?.....	123
1.	LA CONJONCTURE FRANÇAISE DE L'EMERGENCE D'UN CHAMP D'ACTION STRATEGIQUE DE LA METHODE EXPERIMENTALE	127
1.1.	<i>Inscription du travail institutionnel dans l'histoire de l'organisation de la rencontre des méthodes des sciences économiques et de l'Administration centrale en France, à l'écart du champ de l'évaluation des politiques publiques</i>	128
1.1.1.	L'association entre les sciences sociales et l'action publique en France : l'expertise à l'écart des sentiers de l'évaluation	128
1.1.2.	L'expérimentation en France avant la « méthode expérimentale »	136

1.2.	<i>L'étincelle de l'émergence, composition d'un champ embryonnaire : rencontres, organisation et définition des intérêts</i>	145
1.2.1.	La commission « Famille, vulnérabilité et pauvreté », rencontre et amorce d'un projet commun	146
1.2.2.	Les instances d'une réorganisation de l'interaction entre science économique appliquée et action publique autour de l'expérimentation sociale	150
2.	TISSER UNE TOILE D'ALLIANCES AUTOUR D'UN NOUVEAU CHAMP D'ACTION STRATEGIQUE : L'EPREUVE DU POUVOIR (2007-2009)	161
2.1.	<i>La méthode expérimentale, raison d'être, ressource et objectif de l'accession au pouvoir politique</i>	162
2.1.1.	Le cabinet politique comme laboratoire, l'ancrage du champ d'action stratégique dans des champs étatiques	162
2.1.2.	Le RSA, rendez-vous manqué de l'expérimentation, le politique contre la méthode scientifique	165
2.1.3.	L'appel à projets APDIIESES : le difficile contrôle de l'intéressement par le travail cognitif	168
2.2.	<i>Le front scientifique : l'ancrage du champ d'action stratégique dans les champs de l'économie appliquée et de l'expertise d'État dans le secteur de l'emploi</i>	172
2.2.1.	Les expérimentations dans le domaine de l'emploi, l'ancrage dans l'expertise d'État	172
2.2.2.	L'intensification du travail institutionnel pour stabiliser le champ d'action stratégique	179
3.	LA TENTATIVE D'INSTITUTIONNALISATION DU CHAMP D'ACTION STRATEGIQUE DE L'EXPERIMENTATION SOCIALE: STABILISATION ET DISSOLUTION	192
3.1.	<i>Le Fonds d'Expérimentation pour la Jeunesse comme tentative d'institutionnalisation du champ d'action stratégique</i>	193
3.1.1.	La structuration de l'interaction entre science et politique autour de l'unité de gouvernance interne du Fonds d'Expérimentation pour la Jeunesse	194

3.1.2.	Une unité de gouvernance dont les critères scientifiques favorisent les <i>titulaires</i> maîtrisant la ressource « expérimentation aléatoire ».	197
3.1.3.	Une unité de gouvernance interne hybride difficilement contrôlable par les <i>titulaires du champ d'action stratégique</i> : le FEJ « une expérimentation en soi »	199
3.2.	<i>Le détournement d'un champ d'action stratégique par la captation de son unité de gouvernance par les challengeurs</i>	213
3.2.1.	La dispersion du pendant politique du champ d'action stratégique de l'expérimentation sociale.	214
3.2.2.	La subsistance du travail institutionnel cognitif suite à la dissolution du champ d'action stratégique	219
CONCLUSION DU CHAPITRE 2		223
3^{EME} PARTIE : LA METHODE EXPERIMENTALE PAR ASSIGNATION ALEATOIRE ET SES PRATIQUES : ETUDE DES TENSIONS D'APPRENTISSAGES D'UNE RELATION CLINIQUE.....226		
1.	PRESENTATION DES ETUDES DE CAS CONDUITES ET ANALYSEES DANS CE CHAPITRE	230
2.	LA NEGOCIATION DES INTERETS DE LA RELATION CLINIQUE : LA PHASE DE DESIGN	241
2.1.	<i>Les raisons d'être des expérimentations aléatoires, la recherche de l'amélioration des positions d'acteurs au sein de champs en crise</i>	242
2.1.1.	Les évaluateurs/chercheurs, typologie de profils et d'intérêts	242
2.1.2.	Les porteurs de projets expérimentaux, typologie de profils et d'intérêts	251
2.2.	<i>La constitution de la relation clinique</i>	256
2.2.1.	La rencontre des parties prenantes de l'expérimentation	256
2.2.2.	La négociation du design de l'expérimentation	262
3.	L'EPREUVE DU TERRAIN : LA MISE EN ŒUVRE DE L'EXPERIMENTATION ALEATOIRE	267
3.1.	<i>Confrontation des types de travail institutionnel autour de l'assignation aléatoire</i>	268
3.1.1.	Le travail institutionnel de maintien de l'assignation aléatoire	268

3.1.2.	Les facteurs de déstabilisation du protocole expérimental d'assignation aléatoire	272
3.2.	<i>Confrontation des types de travail institutionnel autour du « traitement »</i>	279
3.2.1.	La recherche de la connaissance de la « boîte noire »	279
3.2.2.	Le processus d'apprentissage de la mise en œuvre.....	284
4.	L'UTILISATION DES RESULTATS DE L'EXPERIMENTATION ALEATOIRE PAR SES PARTIES PRENANTES : LA PHASE FINALE DE LA RELATION CLINIQUE	290
4.1.	<i>L'utilisation scientifique des résultats de l'expérimentation aléatoire</i> 290	
4.1.1.	La production des résultats de l'expérimentation aléatoire.....	291
4.1.2.	La capitalisation scientifique issue des expérimentations.....	298
4.2.	<i>Une utilisation des résultats par les acteurs de l'action publique limitée</i>	303
4.2.1.	Les obstacles à l'utilisation des résultats par les acteurs de l'action publique	303
4.2.2.	Bilan de l'apprentissage politique issu des résultats produits	308
	CONCLUSION DU CHAPITRE 3	317
	CONCLUSION	321
	UN INSTRUMENT D'APPRENTISSAGE RENEGOCIE PAR DES INTERACTIONS MULTI-NIVEAUX	322
	<i>Les stratégies d'insitutionnalisation, l'échec de la stabilisation d'un champ d'action stratégique</i>	322
	<i>Une institution d'apprentissage en tension : l'équilibre fragil d'une relation clinique à trois parties</i>	325
	BIBLIOGRAPHIE	329
	TABLE DES FIGURES ET GRAPHIQUES	348
	TABLE DES TABLEAUX ET ENCADRES	349
	TABLE DES ANNEXES	351

ANNEXE 1 : PRESENTATION DU CABINET D’EVALUATION EUREVAL, CO-FINANCEUR DU CONTRAT CIFRE	352
ANNEXE 2 : LISTE DES ENTRETIENS REALISES	353
ANNEXE 3 : ANALYSES DE RESEAU	357
ANNEXE 4 : MEMBRES DU CABINET DU HCSA ET HCJ	362
ANNEXE 5 : TAILLE D’ECHANTILLON ET PUISSANCE STATISTIQUE	363

Remerciements

Je tiens d'abord à remercier vivement tous les interlocuteurs qui se sont rendus disponibles pour s'entretenir avec moi dans le cadre de ma recherche. J'espère que j'aurais réussi à rendre compte de leur passion pour leur travail.

Je souhaite adresser mes profonds remerciements à mon directeur de thèse, Thierry Berthet, pour sa confiance et son soutien dès mes premiers travaux de recherche en master et jusqu'à la finalisation de ce travail de doctorat.

Je remercie aussi infiniment les directeurs d'Euréval, Eric Monnier et Jacques Toulemonde, pour m'avoir ouvert les portes du monde de l'évaluation, et m'avoir soutenue tout en me laissant une liberté totale dans mes travaux de recherche. Leur expertise et enthousiasme sur la pratique évaluative ont été l'impulsion de cette recherche et je ne pourrai jamais assez les en remercier.

C'est d'ailleurs toute l'équipe d'Euréval que je tiens à remercier du fond du cœur pour m'avoir formée avec bienveillance à l'art évaluatif et pour avoir toujours respecté mon travail de recherche. Merci à Christine, Joëlle, Vincent, Virginie, Karine, Hugo, Ximena, Nicolas, Adrien. Un merci tout particulier à Amandine qui a supporté mes doutes existentiels et a illuminé notre bureau de sa légèreté, sa gentillesse, son humour, ses orchidées puis de son amitié. Je remercie aussi particulièrement Thomas Delahais, pour m'avoir stimulée intellectuellement pendant toute cette période. Je le remercie infiniment pour sa disponibilité pour m'accompagner dans l'analyse de réseau, bien que mon utilisation soit très en deçà de la qualité de ses travaux avec cet outil, merci d'avoir transformé mes tableaux abscons en des toiles colorées.

Tous mes remerciements vont aussi vers l'équipe administrative et académique du Centre Emile Durkheim et du Centre Régional Associé du Céreq qui m'ont soutenu tout au long de cette recherche et nourri celle-ci de leur expertise scientifique.

Je tiens à remercier profondément ma première Amandine qui m'a chaleureusement et inconditionnellement accueillie pendant mes séjours à Bordeaux et soutenue dans cette épreuve par toute son amitié. Merci à Stéphanie, Charlène, Nelly, Nadia, Isabelle, Perrine, François, Sylvain, pour leur écoute, leurs encouragements et leur amitié, ils ont été indispensables. Merci à Clemence, de m'avoir montré la voie de cette entreprise doctorale et d'avoir été là, de l'autre côté du monde ou tout près, pour me soutenir jusqu'au bout.

Merci à Amaël pour avoir été mon Ariane dans ce labyrinthe et n'avoir jamais lâché le fil. Ton amour et ton exigence intellectuelle ont été le principal moteur de ce travail.

Enfin, je dédie cette thèse à mes parents et à ma sœur sans qui rien de tout ça n'aurait été possible. Est-ce que le fruit de l'union d'un responsable qualité des traitements pharmaceutiques et d'une travailleuse sociale aurait pu choisir un autre sujet de recherche?

Il va sans dire que toutes les imperfections de ce travail, qu'un lecteur critique ne manquera pas de relever, ne sont la responsabilité que de son auteure.

Lexique des acronymes

AFD : Agence Française de Développement

ANPE : Agence Nationale Pour l'Emploi

ANSA : L'Agence Nouvelle de Solidarité Active

CEREQ : Centre d'Etudes et de Recherches sur les Qualifications

CREDOC : Centre de Recherche pour l'Etude et l'Observation

CREST : Centre de Recherche en Economie et Statistique

DARES : Direction de l'Animation de la Recherche, des Études et des Statistiques

DEPP : Direction de l'Évaluation, de la Prospective et de la Performance

DGEFP : Délégation Générale à l'Emploi et à la Formation Professionnelle

DIIESES : Délégation Interministérielle à l'Innovation à l'Expérimentation Sociale et à l'Économie Sociale

DJEPVA : Direction de la Jeunesse, de l'Education Populaire et de la Vie Associative

DREES : Direction de la Recherche, des Etudes, de l'Evaluation et des Statistiques

ENSAE : Ecole Nationale de la Statistique et de l'Administration Economique

FEJ : Fonds d'expérimentations pour la Jeunesse

INSEE : Institut National des Statistiques et des Etudes Economiques

MAFEJ : Mission d'Animation du Fonds d'Expérimentations pour la Jeunesse

MDRC : Manpower Demonstration Research Corporation

MIT : Massachusetts Institute of Technology

PSE : Paris-School of Economics, Ecole d'Economie de Paris

RSA : Revenu de Solidarité Active

SSM : Services Statistiques Ministériels

TEPP : Fédération Travail Emploi et Politiques Publiques

UNEDIC : Union Nationale Interprofessionnelle pour l'Emploi Dans l'Industrie et le Commerce

Introduction

“Créer une culture qui encouragerait, promouvrait, et financerait des évaluations expérimentales par assignation aléatoire rigoureuses a le potentiel de révolutionner les politiques sociales au XXI^e siècle, comme l’expérimentation par assignation aléatoire a révolutionné la médecine au XX^e siècle »¹.

Ce dessein formulé au début des années 2000 par trois économistes du *Massachusetts Institute of Technology* n’est pas resté lettre morte. Au cours des dix dernières années cette « culture » a connu un développement sans précédent auprès des commanditaires et des prestataires de l’évaluation des politiques publiques. En effet, l’encouragement, la promotion et le financement de la méthode expérimentale par assignation aléatoire devinrent à cette période une entreprise partagée par une diversité d’instances politiques. Au niveau international, la Banque Mondiale en fit sa méthode principale d’évaluation d’impact en moins d’une décennie. Alors que la méthode expérimentale composait seulement 20 % des évaluations d’impact de la Banque mondiale entre 2000 et 2004, elle fut employée pour évaluer l’impact de 83 % des projets initiés entre 2009 et 2010 par cette même institution². Au niveau national, plusieurs ministères du gouvernement fédéral américain l’identifièrent comme la méthode la plus scientifique et favorisèrent sa conduite³. En France, alors qu’elle n’avait jamais été employée pour évaluer des politiques sociales sur le territoire avant 2005, elle fut promue en 2009 par le Haut Commissariat à la Jeunesse comme la méthode à privilégier dans l’évaluation des projets innovants.

Mais en quoi consiste cette méthode d’évaluation ? Et quelle est la nature de la révolution politique dont la méthode expérimentale par assignation aléatoire serait le vecteur ? Cette méthode d’évaluation estime l’effet moyen d’une intervention publique en allouant de manière aléatoire cette intervention à un groupe test et un groupe témoin, puis en comparant l’évolution de chaque groupe vis-à-vis de l’effet souhaité. Sur le modèle du développement de la médecine des preuves (*evidence-based medicine*) par la conduite d’essais cliniques à partir de la deuxième moitié du XX^e siècle, les promoteurs de cette méthode l’envisagent comme l’outil principal de l’instauration d’une nouvelle pratique de l’action publique : *l’evidence-based policy* ; en d’autres termes une construction des politiques publiques s’appuyant davantage sur la prise en compte de travaux scientifiques⁴. La méthode expérimentale par assignation aléatoire employée dans une démarche *d’evidence-based policy* a été investie par ses promoteurs d’une double faculté ; d’une part elle est présentée comme la méthode d’évaluation la plus scientifique pour estimer l’impact d’une intervention, et d’autre part comme la mieux à même d’informer l’action publique et de favoriser l’apprentissage politique. Elle constitue alors un outil ayant la capacité de s’inscrire autant dans les critères d’appréciation d’acteurs scientifiques que ceux des acteurs de l’action publique et d’organiser leur rencontre autour de leur intérêt commun pour les résultats qu’elle produit.

¹ Traduction de l’auteure à partir de l’anglais : “*Creating a culture in which rigorous randomized evaluations are promoted, encouraged, and financed has the potential to revolutionize social policy during the 21st century, just as randomized trials revolutionized medicine during the 20th*” Esther DUFLO, Rachel GLENNERSTER et Michael KREMER, *Randomized Evaluations of Interventions in Social Service Delivery*, MIT.

² INDEPENDANT EVALUATION GROUP, *World Bank Group Impact Evaluations: Relevance and Effectiveness*, Washington, D.C, World Bank, 2012, p. 71.

³ Notamment le ministère de l’Éducation Bernard H MARTIN, *The coalition for evidence-based policy: its impact on policy and practice*, William T. Grant Foundation, 2004.

⁴ Esther DUFLO, « Évaluer l’impact des programmes d’aide au développement: le rôle des évaluations par assignation aléatoire », (2005) 19-2 *Revue d’économie du développement* pp 185–226.

1. L'expérimentation aléatoire, instrument de réconciliation des logiques scientifiques et politiques ?

Cette ambition apparaît paradoxale si elle est mise en perspective avec les différents travaux s'étant intéressés aux pratiques de ces deux sphères ainsi qu'à leur interaction. En effet, l'opposition entre les ambitions, les pratiques et les critères de jugement de l'action publique et de la production scientifique ont été soulignés à plusieurs reprises. Il semble donc, de prime abord étonnant, qu'un même outil puisse organiser une relation médiate entre ces deux univers.

Les logiques du « Savant » et du « Politique » ont été opposées sur plusieurs dimensions. Dans *Le Leviathan*, T. Hobbes oppose ces univers en considérant que la compétition en leur sein est régie par des règles contraires. Alors que les arguments des travaux scientifiques sont départagés par un « *raisonnement solide* » fondé sur les principes de la vérité, les arguments politiques sont régis par la « *puissante éloquence* » fondée sur « *les opinions déjà reçues, vraies ou fausses, et sur les passions et les intérêts des hommes, qui sont divers et changeants* »⁵. En outre, selon Max Weber, ces univers s'inscrivent dans deux éthiques opposées. Le personnel politique fonde son action sur des « *jugements de valeur* » guidés par leur « *éthique de conviction* » [*gesinnungsethisch*] permettant de faire parfois accepter la finalité d'une action quelqu'en soit les moyens, ce que cet auteur estime inconciliable avec « *l'éthique de responsabilité* » [*verantwortungsethisch*] du savant qui ne peut se décharger des conséquences de ses actions⁶. Dans son essai « Vérité et Politique », Hannah Arendt oppose l'activité de recherche de la vérité en science et l'activité politique de rhétorique. Elle explique : « *il n'a jamais fait de doute pour personne que la vérité et la politique sont en assez mauvais termes, et nul autant que je sache n'a jamais compté la bonne foi au nombre des vertus politiques* »⁷. Cette philosophe oppose la logique scientifique de la « *vérité de fait* » qui refuse la discussion, jusqu'à ce qu'elle soit invalidée par une démonstration s'inscrivant dans les mêmes règles de démonstration, à la logique politique constituée, justement sur la discussion et la rhétorique. Elle précise « *l'ennuyeux est que la vérité de fait, comme toute autre vérité, exige péremptoirement d'être reconnue et refuse la discussion alors que la discussion constitue l'essence même de la politique (...). Les modes de pensée et compétitions qui ont affaire avec la vérité, si on les considère dans la perspective politique sont nécessairement tyranniques* »⁸. Cette opposition, d'éthique, de règles de compétition, et d'activité, entre la science et la politique issue d'un travail conceptuel et philosophique de ces trois auteurs a aussi été relevée par des travaux empiriques au cours du XX^e siècle. Nous nous intéressons ici non pas aux travaux sur l'expertise qui proposent d'éclairer la programmation politique par des théories ou des résultats scientifiques produits uniquement dans un cadre scientifique, mais au développement des méthodes et pratiques scientifiques appliquées à l'action publique et conduites dans son contexte et ayant vocation à nourrir sa programmation.

⁵ Thomas HOBBS, *Léviathan: traité de la matière, de la forme et du pouvoir de la république ecclésiastique et civile*, coll. Classiques des sciences sociales., Chicoutimi, J.-M. Tremblay, 2004, p. 73.

⁶ Max WEBER, *Le savant et le politique*, Paris, Plon, 2002, p. 206- 207.

⁷ Hannah ARENDT, « Vérité et politique », dans *La Crise de la culture: huit exercices de pensée politique*, Paris, Gallimard, 1989, p. 289.

⁸ *Ibid.* p. 307.

L'étude de la rencontre de l'action publique et des sciences sociales s'est particulièrement développée au milieu du XX^e siècle avec l'émergence de l'analyse des politiques publiques. En effet, cette nouvelle discipline scientifique est définie par W. Dunn comme « *Une discipline de science sociale appliquée qui utilise des méthodes de recherche multiples dans un contexte d'argumentation de débat public afin de créer, d'évaluer de manière critique, et de communiquer une connaissance pertinente pour l'action publique* »⁹. Ainsi, par cette discipline, des chercheurs se sont davantage investis dans la pratique de l'action publique, créant de nouvelles connaissances sur son fonctionnement ainsi que sur la nature de son interaction avec la science. À partir du développement de ces pratiques évaluatives, particulièrement dans les pays anglo-saxons, des chercheurs, observateurs extérieurs ou s'étant engagés eux-même dans une entreprise évaluative, ont questionné la pertinence de transférer des outils scientifiques dans le champ de l'action publique. Après la Seconde Guerre mondiale jusque dans les années 1980, l'analyse des politiques publiques d'abord envisagée comme un outil de conseil au service de la démocratie par A. Laswell, devient un outil de production scientifique et ayant vocation à appliquer son protocole de production de connaissances à l'action publique, notamment par les premières utilisations de la méthode expérimentale par assignation aléatoire¹⁰.

À partir des années 1980, s'opère un tournant « post positiviste » questionnant la pertinence de l'utilisation de données produites par des méthodes scientifiques pour nourrir l'action publique. C. Lindblom explique que la conduite de l'action publique s'écarte d'une logique de résolution de problèmes publics cherchant l'intervention la plus efficace. Il en déduit que l'impact du chercheur est d'une portée réduite et se limite à nourrir l'argumentaire des luttes rhétoriques entre politiciens¹¹. A. Wildavsky critique le manque de réalisme politique des méthodes scientifiques et souligne leur habilité à véhiculer aussi des théories idéologiques et politiques sous couvert de technicité¹². Ces critiques ouvrent au développement d'un nouveau type d'association des sciences sociales et du politique se concentrant avant tout sur la compréhension du fonctionnement de l'action publique, la prise en compte de la diversité d'intérêt, et une appréciation des méthodes employées non sur leur scientificité, mais sur leur utilité pour éclairer la décision¹³. Les travaux scientifiques mobilisés pour éclairer l'action publique sont considérés non pas comme supérieurs aux autres arguments, mais comme un ensemble d'arguments qu'il faut considérer comme tel et contribuant au pluralisme politique¹⁴.

⁹ Traduction de l'auteure à partir de l'anglais : « *applied social science discipline that uses multiple research methods in a context of argumentation, public debate in order to create, critically evaluate, and communicate policy-relevant knowledge* » William N DUNN, *Public policy analysis: an introduction*, Boston, Pearson, 1994, p. 84.

¹⁰ Robert HOPPE, « Policy analysis, science, and politics: from "speaking truth to power" to "making sense together" », (1999) 26-3 *Science and Public Policy* 201-210.

¹¹ Charles Edward LINDBLOM et David K COHEN, *Usable knowledge: social science and social problem solving*, New Haven, Yale University Press, 1979; Charles LINDBLOM, « The Science of "Muddling Through" », (1959) 19 *Public Administration Review* 79-88.

¹² Aaron B WILDAVSKY, *Speaking truth to power : the art and craft of policy analysis*, New Brunswick, 1987.

¹³ Carol H WEISS, « Evaluation Research for Decision: Is Anybody There? Does Anybody Care? », dans Marvin C ALKIN, *Debates on evaluation*, Newbury Park, Calif., Sage Publications, 1990; Michael Quinn PATTON, *Utilization-focused evaluation : the new century text*, Thousand Oaks; London; New Delhi, Sage Publications, 1997.

¹⁴ Dennis James PALUMBO (dir.), *The Politics of program evaluation*, coll. Sage yearbooks in politics and public policy, Newbury Park, Calif, Sage Publications, 1987; John S. DRYZEK, « Policy Analysis and Planning: From Science to Argument », dans Frank FISCHER et John FORESTER, *The Argumentative Turn in Policy Analysis and Planning*, Taylor & Francis, 1993, p. 213-232.

Nous identifions deux constats particulièrement intéressants pour notre objet de recherche dans ces travaux.

D'abord, une distinction de pratique se dessine entre les scientifiques « appliqués » et les autres, présentant les premiers comme s'inscrivant aussi en partie dans des logiques argumentatives correspondant aux normes et standards de la compétition politique. Cette particularité s'incarne aux États-Unis, ainsi qu'en Europe par l'émergence de formations et de professions de chercheurs appliqués, comme les statisticiens de l'INSEE, ou les évaluateurs qui par leur investissement dans l'action publique se distinguent et s'écartent de la logique et de la compétition scientifique traditionnelle. Comme le soulignait A. Desrosières, ces nouvelles sciences appliquées comme notamment la statistique apparaissent d'avantage comme résultant d'une co-construction de l'État et de la science¹⁵. Ces instruments d'apparence neutre et techniques sont donc aussi porteurs d'intérêts et de théories, ce qui ouvre la possibilité de l'étude sociologique de ces derniers. Ainsi, il semble que pour appréhender la rencontre de la science et de la politique, il est de nos jours davantage pertinent de concevoir les caractéristiques de leurs logiques, éthiques et règles de compétitions particulières, plutôt comme deux extrémités d'un continuum sur lequel se place une variété de pratiques présentant une combinaison de ces caractéristiques.

Un second constat, est que malgré cette nouvelle proximité de pratiques et les efforts des sciences sociales pour s'adapter au champ politique, leur influence reste très limitée et soumise aux intérêts politiques. Des travaux récents soulignent encore le manque de prise en compte des résultats d'évaluation dans la décision publique¹⁶. Jusqu'en 2008, E. Chelimsky ancienne directrice du service de méthodes et d'évaluation du *General Accountability Office* (GAO) - équivalent américain de la Cour des comptes – évoquait un « *Clash des cultures* » entre ces deux sphères où le constat scientifique peinait à prendre sa place dans un contexte politique¹⁷.

Au vu de ces travaux, l'ambition de la méthode expérimentale par assignation aléatoire de se positionner comme un Janus, une porte, ouvrant une communication réconciliée entre d'une part, une production scientifique s'inscrivant dans ses standards les plus exigeants, et d'autre part, la décision politique qui s'orienterait vers un recours systématiquement de la preuve (*evidence-based policy*), en fait effectivement un outil « révolutionnaire » dont les modalités de fonctionnement appellent à être étudiées.

Ce travail de recherche est articulé autour d'une problématique générale questionnant en quoi et de quelle manière la méthode expérimentale par assignation aléatoire appliquée à l'action publique constitue un instrument réorganisant les rapports entre les sciences sociales et l'action publique.

Cette problématique s'inscrit dans le programme de recherche ouvert par l'étude *des instruments d'action publique* que nous avons suivi dans ce travail en développant un appareillage théorique néo-institutionnaliste sociologique permettant de rendre compte de l'émergence de cette méthode et de son impact sur son environnement.

¹⁵ Alain DESROSIÈRES, *Statistique et espace public: le rêve de l'extériorité, Le savant et le politique aujourd'hui*, Bibliothèque Albin Michel, Colloque de La Villette, 1996, p. 107-118.

¹⁶ Carol H. WEISS, « The Fairy Godmother and Her Warts: Making the Dream of Evidence-Based Policy Come True », (2008) 29-29 *American Journal of Evaluation*; Michael Quinn PATTON, *Developmental evaluation : applying complexity concepts to enhance innovation and use*, New York, Guilford Press, 2010.

¹⁷ Eleanor CHELIMSKY, « A Clash of Cultures: Improving the "fit" Between Evaluative Independence and the Political Requirements of a Democratic Society », (2008) 29-4 *American Journal of Evaluation* 400-415.

2. Un cadre néo-institutionnaliste sociologique pour appréhender une méthode scientifique et ses effets

Les travaux de la sociologie de l'action publique se sont intéressés à des initiatives de rationalisation, et de technicisation de l'action publique par l'application de méthodes organisant l'action et l'apprentissage politique et se regroupent autour de la notion d'*instrument d'action publique (IAP)*. L'évaluation des politiques publiques a déjà été caractérisée comme un instrument d'action publique¹⁸. Néanmoins, sa diversité d'incarnation rendait cet objet peu préhensible pour le chercheur, alors que de la méthode expérimentale par assignation aléatoire en définit un périmètre beaucoup plus strict et maîtrisable. Selon C. Hood, trois approches se distinguent dans l'étude des instruments. La première s'intéresse aux choix des instruments et l'adéquation entre l'objectif et les moyens d'une politique publique, la seconde porte sur la diversification et la multiplication des instruments, la dernière appréhende les instruments comme des institutions sociologiques¹⁹. Cette recherche relève de la troisième démarche. Nous considérons comme la première citée que les instruments sont des variables dépendantes dont il convient d'étudier les processus sociopolitiques d'émergence et de diffusion et de les appréhender en tant qu'unité d'observation indépendamment des secteurs auxquels ils s'appliquent²⁰. Cependant nous les étudions comme des institutions sociales au sens néo-institutionnaliste c'est-à-dire en prenant en compte leur dimension symbolique et cognitive²¹. Cette démarche porte d'une part, sur la genèse des instruments en les considérant comme sujets à des renégociations et évolutifs au gré des contextes et d'autre part, sur leurs effets quant aux représentations et aux pratiques des acteurs auxquels ils s'appliquent. Nous proposons de nous intéresser dans ce travail à cette double dynamique en mobilisant les travaux théoriques récents sur la *théorie générale des champs* et le *travail institutionnel*²². Nous développons plus en détail ces choix théoriques dans le chapitre introductif de ce travail.

Le courant sociologique de l'étude des IAP définit ces derniers comme « *un dispositif normatif, à la fois technique et social, ayant une vocation générique (applicable à un ensemble de situations) et porteur d'une conception concrète du rapport gouvernant/gouverné fondé sur une conception spécifique de la régulation* »²³. L'évaluation des politiques publiques correspond à cette définition, mais jusqu'à présent ses diversités d'incarnations en compliquaient l'analyse comme un objet homogène, défendant une conception univoque du rapport gouvernant/gouverné et de la régulation. En effet, les différents travaux s'étant intéressés à cet objet, ont souligné les controverses méthodologiques la structurant ainsi que leurs différentes vocations sociétales²⁴. En outre, notre immersion dans le milieu des évaluateurs et la pratique de

¹⁸ Frédéric VARONE et Steve JACOB, « Cheminement institutionnel de l'évaluation des politiques publiques en France, en Suisse et aux Pays-Bas (1970-2003) », (2004) 22-2 *pomap* 135-152.

¹⁹ Pierre LASCOUMES et Louis SIMARD, « L'action publique au prisme de ses instruments », (2011) 61-1 *Revue française de science politique* 5-22, p. 11.

²⁰ Frédéric VARONE, « Le choix des instruments de l'action publique: analyse comparée des politiques énergétiques en Europe et Amérique du nord », (2000) 7-1 *Revue Internationale de Politique Comparée* 167-201.

²¹ Pierre LASCOUMES et Louis SIMARD, « L'action publique au prisme de ses instruments », (2011) 61-1 *Revue française de science politique* 5-22, p. 17.

²² Neil FLIGSTEIN et Doug MCADAM, *A theory of fields*, New York, Oxford University Press, 2012; Thomas B. LAWRENCE et Roy SUDDABY, « Institutions and Institutional Work », dans Stewart CLEGG, C HARDY et T.B. LAWRENCE, *The Sage handbook of organization studies*, 2nd ed, London ; Thousand Oaks, Calif, Sage Publications, 2006, p. 215-254.

²³ Pierre LASCOUMES et Louis SIMARD, *op. cit.* p. 18.

²⁴ L'opposition la plus saillante est notamment celle en France entre le modèle managérial de l'évaluation où cette dernière est employée pour estimer la valeur numérique de l'atteinte des objectifs de l'action publique et le modèle pluraliste où l'évaluation est employée comme outil démocratique de délibération et de construction d'un consensus de valeur E GUBA et Y LINCOLN,

ce métier pendant plus de trois ans, nous a montré que même parmi des évaluateurs s'inscrivant dans une même démarche évaluative, les méthodes mobilisées sont toujours constituées de manière *ad hoc* par un mélange complexe d'outils qualitatifs, quantitatifs, ou de communication s'adaptant aux secteurs d'interventions, aux commanditaires, et aux questions évaluatives. Comme le soulignent des observateurs américains « *l'évaluation - plus que n'importe quelle science - est ce que les gens disent qu'elle est, et les gens sont en train de dire que ce peut être beaucoup de choses différentes* »²⁵. Cette diversité composait donc un obstacle à la définition d'un objet d'étude préhensible dont on pourrait étudier les effets de manière rigoureuse.

La méthode expérimentale par assignation aléatoire fait rupture avec ces différentes démarches évaluatives puisque ses promoteurs la présentent comme linéaire, transparente, systématique et standardisée²⁶. Cette méthode propose d'évaluer un dispositif public, en allouant ses bénéficiaires potentiels par assignation aléatoire à un groupe test — bénéficiant de l'intervention évaluée — et un groupe témoin — n'en bénéficiant pas — et d'en estimer l'effet, en calculant le différentiel entre les deux groupes selon les indicateurs d'impact sélectionnés. Elle se présente donc, de prime abord, comme un objet dont on peut clairement délimiter les contours et stable dans sa diffusion. De plus, elle est aussi présentée par ses promoteurs comme l'instrument d'une conception particulière de régulation et de rapport gouvernant/gouverné : *l'evidence-based policy*. La clarification de cette notion étant une partie intégrante de notre recherche nous reviendrons sur celle-ci dans notre premier chapitre. Cette méthode a aussi un usage générique, elle s'emploie dans des disciplines scientifiques, en psychologie, économie ou agronomie par exemple, et peut s'appliquer à plusieurs secteurs d'activité dans le cadre de pratiques *evidence-based* comme *l'evidence based medicine* ou *l'evidence-based education*. Nous concentrons cette étude sur son application à l'action publique lorsqu'elle est inscrite dans une démarche *d'evidence-based policy*.

Ce programme de recherche implique d'étudier les IAP comme des institutions sociales et de fait implique de s'inscrire dans les travaux néo-institutionnalistes s'intéressant à l'émergence des institutions et à leurs effets. Cependant, au sein du courant du néo-institutionnalisme, penser l'institution s'opère de manière variée et un positionnement parmi ces travaux s'impose. Nous revenons dans le chapitre 1 plus en détail sur les travaux s'étant intéressés d'une part, à l'émergence d'institutions (Chapitre 1, partie 2.1) et d'autre part, à leur influence sur les acteurs qui les composent (Chapitre 1, partie 2.2). Nous présentons ici, brièvement, notre manière d'envisager l'institution tout au long de ce travail. Cette recherche est traversée par une conception constructiviste de la relation entre l'institution et l'acteur s'inscrivant dans la théorie de la structuration d'A. Giddens²⁷. En d'autres termes, nous considérons que si des institutions peuvent cadrer les possibles des interactions sociales, elles n'existent néanmoins que par les pratiques et les interprétations des acteurs qui peuvent en retour les modifier. Nous sommes donc

« The countenances of 4 generation evaluation : Description, judgement and negociation », dans Dennis James PALUMBO, *The Politics of program evaluation*, coll. Sage yearbooks in politics and public policy, Sage Publications, 1987, p. 202-234; Patrice DURAN et Eric MONNIER, « Le développement de l'évaluation en France. Nécessités techniques et exigences politiques », (1992) 42-2 *revue française de science politique* 235-262.

²⁵ Traduction de l'auteure à partir de l'anglais: « *Evaluation -more than any science- is what people say it is; and people currently are saying it is many different things* » William R SHADISH, Thomas D COOK et Laura C LEVITON, « Social Program Evaluation: Its History, Tasks, and Theory », dans *Foundations of program evaluation: theories of practice*, Sage Publications, 1991, p. 19-35 p. 30.

²⁶ Esther DUFLO, « Evaluer l'impact des programmes d'aide au développement: le rôle des évaluations par assignation aléatoire », *op. cit.*

²⁷ Anthony GIDDENS, *La constitution de la société: éléments de la théorie de la structuration*, Paris, Presses universitaires de France, 2012.

face à une dynamique réursive, où les structures sociales sont à la fois constitutives de l'interaction des acteurs et constituées par eux. Comme le soulignait E. Goffman, nous sommes en présence de deux types d'ajustements, un premier cadrant les possibles, et un second ouvrant des capacités d'adaptation des acteurs pouvant entraîner la modification du premier ajustement²⁸. Cette conception de l'institution se retrouve dans l'appareillage pour concevoir l'émergence d'institutions ainsi que leur impact sur leur environnement.

Selon cette perspective les institutions n'émergent, ni uniquement par l'agence d'un groupe d'acteurs, ni seulement par des facteurs conjoncturels déterminants. Comme le souligne A. Giddens, en évoquant la construction de l'Etat, les institutions émergent grâce, d'une part à une conjoncture définie comme « *des réseaux d'influence qui, à un moment et à un endroit précis, sont pertinents pour un épisode particulier* » rencontrant d'autre part, l'influence de la compétence humaine sur le changement social²⁹. Ainsi pour étudier la sociogenèse d'une institution, il est nécessaire d'abord de comprendre la conjoncture de son émergence et ensuite les moteurs de l'influence des acteurs et leurs modalités d'actions. Afin de comprendre les réseaux d'influence, nous avons mobilisé le concept de Champ d'Action Stratégique (*strategic action field*) défini par N. Fligstein et D. McAdam comme « *ordre social construit, de niveau intermédiaire, au sein duquel les acteurs (qui peuvent être individuels ou collectifs) ont conscience les uns des autres et interagissent sur la base d'une compréhension partagée (ce qui ne veut pas dire consensuelle) sur les buts du champ, les relations entre les personnes du champ (dont qui a le pouvoir et pourquoi), ainsi que les règles gouvernant l'action légitime dans le champ* »³⁰. Les champs sont ici conçus comme des poupées gigognes de taille variables s'imbriquant les uns dans les autres, ou pouvant se chevaucher et s'influençant les uns les autres. Leurs contours, leurs enjeux et la hiérarchie en leur sein se modifient au gré des conjonctures et des interactions entre les acteurs les composant cherchant à améliorer leur position. Néanmoins, ces acteurs en retour, subissent aussi l'influence des champs qui organisent leur *pratiques* propres à leur champ d'appartenance qui peuvent, selon les conjonctures, les doter de compétences plus ou moins pertinentes pour agir dans un contexte donné. En ayant recours au concept de *travail institutionnel*, nous considérons que cette entreprise de construction institutionnelle émergente est un processus mouvant et s'ajustant aux différentes interactions qui la structure et au travail de maintien ou de déstabilisation que les entrepreneurs institutionnels rencontrent dans leur entreprise. Ce concept nous amène donc à nous intéresser à un cercle élargi d'acteurs dans l'étude d'une entreprise de construction institutionnelle.

Dans un deuxième temps, nous nous intéressons aux effets d'une institution sur les acteurs qui la composent et son environnement. Une perspective constructiviste implique deux points d'attention dans cette entreprise. Elle nécessite d'appréhender l'institution comme susceptible de s'incarner dans une variété de *sites institutionnels* pour lesquels il convient, comme le souligne O. Nay, de « *rendre compte de certaines logiques sociales spécifiques qui sont propres à l'institution politique [et scientifique pour notre objet] observée, et d'autre part à saisir la pluralité des usages et des représentations que ces logiques induisent dans des contextes d'interaction particuliers*

²⁸ Erving GOFFMAN, *Asiles; études sur la condition sociale des malades mentaux et autres reclus*, Paris, Editions de minuit, 1968.

²⁹ Anthony GIDDENS, *op. cit.* p. 312.

³⁰ Traduction de l'auteur de l'anglais "A strategic action field is a constructed mesolevel social order in which actors (who can be individual or collective) are attuned to and interact with one another on the basis of shared (which is not to say consensual) understandings about the purposes of the field, relationships to others in the field (including who has power and why), and the rules governing legitimate action in the field" Neil FLIGSTEIN et Doug MCADAM, *A theory of fields*, New York, Oxford University Press, 2012, p. 9.

mettant aux prises des acteurs porteurs de rationalités et d'exigences différenciées.»³¹. Afin de rendre compte et d'expliquer les différentes incarnations institutionnelles, nous mobilisons le concept de travail institutionnel (*institutional work*), élaboré par T. Lawrence et R. Suddaby, qui désigne « l'action intentionnelle des acteurs ou des organisations visant à créer, maintenir ou déstabiliser les institutions »³². Nous identifierons les différents types de travaux mis en place et leur degré variable d'intentionnalité. A titre d'exemple, un travail de déstabilisation peut juste être issu d'une *pratique* professionnelle routinisée allant à l'encontre de cette nouvelle institution ou être provoqué par un travail actif intentionnel des acteurs. Nous chercherons à expliquer les motivations de ce travail institutionnel en mobilisant à nouveau la théorie générale des champs. Nous identifierons les champs d'action stratégiques d'appartenance de ces différents acteurs et caractériserons leur position dans ces derniers afin de dégager des facteurs explicatifs du travail institutionnel de maintien ou de déstabilisation dans lequel ils s'engagent. Dans un deuxième temps, par une analyse croisée des différents sites institutionnels étudiés, nous dégagerons les « *logiques sociales spécifiques* » propres à l'institution étudiée, en d'autres termes, les cadres et les théories d'action dont elle est porteuse. La méthode expérimentale constituant une méthode d'apprentissage scientifique et ayant vocation par son utilisation dans le cadre de *l'evidence-based policy* à favoriser l'apprentissage politique, nous nous questionnerons particulièrement sur les modalités d'apprentissages dont elle est porteuse. Nous reviendrons en détail sur cette question de l'apprentissage dans le chapitre 3 de ce travail.

Afin de conduire l'étude de cet instrument de manière empirique, il est nécessaire de délimiter un terrain de recherche restreint de son application. Néanmoins, cet instrument étant importé d'un autre contexte d'application, un retour sur son origine est un passage obligé pour en comprendre les enjeux. Nous avons donc choisi de procéder par une première partie introductive pour restituer au lecteur la sociogenèse de la méthode expérimentale par assignation aléatoire, en nous appuyant principalement sur des données secondaires et permettant de situer l'environnement et le contexte international de notre cas. L'étude empirique de cette recherche se concentre sur le territoire français. Nous procédons ici à une démarche informelle d'études de cas emboîtées (*embedded case study*), au sens où la première unité de cas est le contexte national français et l'étude de la diffusion de cette méthode à un niveau meso-sociologique. La deuxième unité de cas, se situe à un niveau microsociologique, l'expérimentation, elle-même « emboîtée » dans cette première unité³³. Nous étudions donc, un cas national et 15 cas d'expérimentations dans ce contexte.

³¹ Cet auteur définit les sites institutionnels comme « des espaces sociaux identifiés par trois critères : un ensemble d'acteurs interdépendants (configuration) ; l'existence d'interactions menées en continu (dynamique de jeux) ; et enfin la présence de règles institutionnalisées exprimées dans un ensemble de manières d'agir et de penser (niveau microsociologique) et de logiques sociales (niveau macrosociologiques) pesant sur l'ensemble des comportements et des usages des acteurs en situation de co-présence ». Olivier NAY, *La Région, une institution: la représentation, le pouvoir et la règle dans l'espace régional*, Paris, Harmattan, 1997, p. 20.

³² Thomas B. LAWRENCE et Roy SUDDABY, « Institutions and Institutional Work », dans Stewart CLEGG, C HARDY et T.B. LAWRENCE, *The Sage handbook of organization studies*, Sage Publications, 2006, p. 215-254 p. 215., Nous reprenons ici la traduction de l'anglais : « the purposive action of individuals and organizations aimed at creating, maintaining and disrupting institutions » proposée dans, Ben Slimane, Karim, et Bernard Leca. 2010. « Le travail institutionnel : origines théoriques, défis et perspectives ». *Management & Avenir* 7 (37): 53-69,

³³ Roland W. SCHOLZ et Olaf TIETJE, *Embedded case study methods: integrating quantitative and qualitative knowledge*, Thousand Oaks, Calif, Sage Publications, 2002.

Au-delà de sa facilité d'accès dans le cadre de cette recherche, le cas français est particulièrement intéressant pour étudier cette problématique à plusieurs égards. De prime abord, la France apparaît comme un terrain vierge de l'emploi de l'expérimentation aléatoire pour évaluer les programmes sociaux avant 2005. Étudier ce cas, permet donc d'apprécier pleinement, dans une histoire récente, les acteurs de son développement et son impact sur la transformation des relations entre sciences sociales appliquées et action publique. Cet effet par exemple, est plus difficilement appréciable dans des pays comme les États-Unis où cette méthode fut utilisée sur des projets de grande ampleur dès les années 1960 et où elle a déjà contribué à forger les modalités d'interactions entre ces deux sphères. De plus, le développement de cette méthode a été très rapide en France, puisqu'elle fut désignée par une instance de l'administration centrale en 2009, comme la méthode à privilégier pour l'évaluation de projets innovants financés par des appels à projets³⁴. Enfin, ce développement rapide soulève plusieurs questionnements pour les observateurs du développement de l'évaluation des politiques publiques en France au sein de l'administration centrale. En effet, les travaux sur l'institutionnalisation des pratiques évaluatives en France, soulignent unanimement que, malgré les initiatives successives, l'évaluation conduite par des acteurs externes est restée une pratique très marginale au sein de l'administration centrale³⁵. La pratique d'évaluation était principalement limitée à la Cour des comptes et aux Services Statistiques Ministériels (SSM) tels que la Dares ou la Depp³⁶. Or, la méthode expérimentale par assignation aléatoire a été présentée justement par les institutions y ayant recours comme une méthode organisant une évaluation externe et indépendante des interventions du fait qu'elle soit pratiquée par des acteurs scientifiques. Le recours à cette méthode et son financement marquerait-il un renouveau dans l'institutionnalisation de l'évaluation au sein de l'administration centrale française ?

3. Une stratégie de recherche itérative entre collecte de données et théorie

Afin de nourrir ces questionnements, nous avons dans ce travail élaboré une stratégie de recherche articulant travail empirique et théorique par abduction. Cette dynamique itérative entre théorie et observation qui a guidé notre collecte et l'analyse des données résulte en partie du contexte de conduite de cette thèse dans le cadre d'un contrat Cifre (Conventions Industrielles de Formation par la REcherche).

3.1. Le contrat Cifre, de la participation observante à l'observation participante

Ce travail de doctorat a été conduit dans le cadre d'un contrat Cifre (Conventions Industrielles de Formation par la REcherche) de trois ans – d'octobre 2010 à octobre 2013 - avec le cabinet d'évaluation de politiques publiques Euréval (voir présentation en annexe1). Ce choix est issu d'une part de recherche de financement et d'autre part de la nécessité de comprendre le monde évaluatif français, ses systèmes de représentations et ses pratiques. Nous n'avons pas envisagé l'observation participante et l'étude ethnographique comme une méthode centrale de notre recherche. Néanmoins, ce type de

³⁴ CONSEIL SCIENTIFIQUE DU FEJ, *Guide méthodologique pour l'évaluation des expérimentations sociales à l'intention des porteurs de projets*, 2009.

³⁵ Frédéric VARONE et Steve JACOB, *op. cit.*; Vincent SPENLEHAUER, *L'évaluation des politiques publiques, avatar de la planification*, Grenoble, Université Pierre Mendès-France, 1998; Patrice DURAN et Eric MONNIER, *op. cit.*

³⁶ La DARES (Direction de l'Animation de la Recherche, des Études et des Statistiques) est le SSM du ministère du Travail et la DEPP (Direction de l'Évaluation, de la Prospective et de la Performance) celui du ministère de l'éducation. Il existe 16 Services Statistiques Ministériels leur liste est en annexe du *Décret n° 2009-250 du 3 mars 2009 relatif à l'Autorité de la statistique publique*, (2009).

financement a eu une influence déterminante sur l'ensemble de ce travail, bien que nous n'ayons pas choisi comme terrain de recherche le cadre de notre activité professionnelle. Les activités professionnelles et les activités de recherche ont été menées de façon distincte. Cependant, cette immersion dans le milieu de l'évaluation a facilité considérablement notre accès aux données, notamment dans le cadre de notre engagement associatif à la Société Française d'Evaluation. Nous proposons ici, de caractériser ce positionnement et la place qu'il a joué dans la conduite de ces recherches, à mi-chemin entre l'observation participante et de la participation observante. Notre démarche ne correspond pas à de l'observation participante selon la définition d'usage³⁷. Nous avons certes été immergée pendant une période prolongée dans l'environnement professionnel d'un cabinet d'évaluation et partagé le quotidien des évaluateurs ; Cependant, cette situation n'a pas nourri une collecte systématique de données sur cette expérience professionnelle. L'immersion dans ce milieu a été mobilisée de différentes manières selon les moments de la recherche.

Dans un premier temps, elle a été mobilisée dans un souci d'aborder de façon inductive le milieu de l'évaluation, afin de définir un objet de recherche en connaissance du contexte. D'une certaine manière, ce contrat a été un moyen de négocier l'accès au terrain ou « l'entrée » dans le milieu de l'évaluation³⁸. Le manque patent d'études existantes en sciences sociales sur l'évaluation et notamment sur sa pratique avait été un handicap pour appréhender ce phénomène lors d'un premier travail de recherche. Souhaitant continuer à travailler sur l'évaluation dans le cadre d'un travail de thèse, une part d'observation participante dans ce milieu semblait essentielle pour comprendre les enjeux de ce dernier, son vocabulaire de travail et son interaction avec le monde politique. Le choix du cabinet d'accueil n'a pas été au hasard. Ce cabinet d'évaluation de politiques publiques est le plus ancien en France – il fut créé en 1994 — et ses directeurs sont des acteurs influents du monde de l'évaluation. Ils ont élaboré les standards d'évaluation de l'Union européenne, fondé le premier master d'évaluation des politiques publiques à l'Institut d'Etudes Politiques de Lyon et participé dans le cadre du Conseil Scientifique de l'Evaluation aux débats méthodologiques autour des tentatives d'institutionnalisations de l'évaluation au sein de l'administration centrale³⁹. L'élaboration du sujet de thèse et le début du doctorat ont été précédés par un contrat de travail à durée déterminée de six mois avec l'entreprise d'accueil. Ce contrat fonctionnait comme une période d'essai pendant laquelle l'entreprise jugeait de notre « compatibilité » avec le monde de l'entreprise. En contrepartie, cette période de six mois était aménagée pour que nous puissions consacrer du temps à l'élaboration d'un sujet de thèse. Cette période a d'abord été un moment de formation à l'évaluation. Ce fut aussi un moment de lecture intensive de la littérature professionnelle et scientifique de l'évaluation via les abonnements de l'entreprise aux revues anglo-saxonnes faisant référence dans le domaine (*Evaluation*, *Evaluation Review*, *New direction for evaluation*). Le choix de ce sujet est issu d'une approche majoritairement inductive, puisqu'il résulte de la lecture de la littérature professionnelle et scientifique dans ce

³⁷ Stéphane DUFOR, Dominic FORTIN et Jacques HAMEL, *L'enquête de terrain en sciences sociales : l'approche monographique et les méthodes qualitatives : bibliographie annotée*, Montréal, Éditions Saint-Martin, 1991, p. 25; Steven J TAYLOR et Robert BOGDAN, *Introduction to qualitative research methods : a guidebook and resource*, New York, Wiley, 1998.

³⁸ George LAPASSADE, « Observation participante », dans Eugène ENRIQUEZ, Jacqueline BARUS-MICHEL et André LEVY, *Vocabulaire de psychosociologie*, Paris, [diffusion] Cairn.info, 2010, p. 375-390.

³⁹ La première mission de ce Cabinet a été la rédaction des guides MEANS (Méthodes d'évaluation d'actions de nature structurelle) établissant les standards d'évaluation européens. E. Monnier était membre du Conseil Scientifique de l'Évaluation et il fait partie avec Jacques Toulemonde (autre directeur d'Euréal) des membres fondateurs de la Société Française d'Évaluation. UNION EUROPEENNE. DIRECTORATE-GENERAL FOR REGIONAL POLICY AND COHESION, *Evaluer les programmes socio-économiques : Principales techniques et outils d'évaluation*, 3, coll. MEANS, Luxembourg, Office des publications officielles des Communautés européennes, 1999; CONSEIL SCIENTIFIQUE DE L'ÉVALUATION, *Petit guide de l'évaluation des politiques publiques*, Paris, La Documentation française, 1996.

domaine, de sa mise en perspective avec le contexte d'immersion, et de leur rencontre avec nos connaissances en sociologie de l'action publique.

Nous avons délibérément souhaité que cette recherche ne porte pas sur l'environnement quotidien direct d'immersion. La distanciation nécessaire face à l'objet, ainsi que la volonté de ne pas avoir un impact trop important sur le terrain étaient difficilement envisageables dans une situation de contact quotidien avec l'objet pendant une période de trois ans. La pratique de l'activité professionnelle se rapprochait d'une approche de participation observante. Au sens où la pratique de l'activité professionnelle et l'implication interactionnelle et intersubjective prenaient le dessus sur l'observation⁴⁰. En d'autres termes, comme le souligne L. Wacquant, dans un premier temps, nous avons appris davantage sur le milieu de l'évaluation par l'apprentissage de cette posture professionnelle que par son observation⁴¹. Cet accès à la pratique professionnelle nous a donné des clefs de compréhension des interactions entre les différents acteurs de ce milieu ainsi que de leur choix de positionnement. Cet apprentissage au cœur de ce milieu, nous a permis d'acquérir une reconnaissance des acteurs français de l'évaluation comme l'une des leurs, appartenance amplifiée par le titre de co-animatrice du réseau des chercheurs en évaluation des politiques publiques de la Société Française d'Évaluation.

Nous avons pris la précaution de choisir un objet à la périphérie de notre activité professionnelle et associative, puisque le monde de l'évaluation expérimentale est à part du monde de l'évaluation et non représenté au sein de la Société Française d'Évaluation. Nous avançons à visage semi-découvert, au sens où nos interlocuteurs avaient connaissance de notre double positionnement de chercheuse et professionnelle, notre objet d'étude était connu de la plupart d'entre eux, mais sans pour autant que ces derniers aient connaissance de nos hypothèses de recherche. Le statut acquis au sein de cette association, nous donnait accès aux réflexions informelles des figures de proue de l'évaluation avec lesquelles nous étions en contact, sur le développement de la méthode expérimentale. L'apprentissage de l'évaluation « à la française » dans les règles de l'art - notre cabinet d'accueil étant le plus important formateur professionnel et universitaire en évaluation – nous fournissait les clefs pour comprendre et analyser les différents positionnements par rapport à cette nouvelle méthode d'évaluation.

Néanmoins, nous avons découvert lors de notre phase de terrain que l'appartenance à cette association constituait un marqueur d'opposition à la méthode expérimentale par assignation aléatoire. Notre positionnement associatif et professionnel devenait alors à double tranchant. D'une part, il était un atout majeur pour l'accès aux interlocuteurs, armée du vocabulaire professionnel technique pour mener nos entretiens et des clefs pour interpréter les sous-entendus et les discours de chacun. D'autre part, il avait une influence sur notre collecte de données dans le cadre de la conduite d'entretiens avec les praticiens de la méthode expérimentale qui pouvaient nous avoir identifiée comme potentielle opposante. Bien que nous ayons eu pour souci de défendre la neutralité axiologique de notre travail auprès de nos interlocuteurs, cette dimension est à prendre en compte dans l'analyse du matériau collecté. Quoi qu'il en soit, les avantages que nous avons retiré de cette immersion vis-à-vis notamment de notre compréhension du milieu de l'évaluation et de l'impact de notre objet sur le *statu quo* évaluatif français, surpassent les difficultés dont elle est la source. L'observation cependant, n'est par la

⁴⁰ Bastien SOULE, « Observation participante ou participation observante? Usages et justifications de la notion de participation observante en sciences sociales », (2007) 27-1 *Recherches Qualitatives* 127- 140, 130.

⁴¹ Loïc WACQUANT, *Corps & âme : carnets ethnographiques d'un apprenti boxeur*, Marseille, Agone, 2002, p. 11.

méthode principale de notre recherche puisque nous avons choisi de croiser l'apport de différents outils de collecte et d'analyse.

3.2. Appareillage de collecte de données

Ainsi le découpage de cet objet de recherche et sa problématique sont issus d'une rencontre entre les théories de la sociologie de l'action publique et notre immersion dans la communauté des évaluateurs français et la pratique de cette activité. Notre méthodologie de collecte de données est aussi empreinte de ces deux disciplines. Nous développons une stratégie de collecte et d'analyse de données différente dans chaque chapitre puisque nous l'adaptions aux diverses questions abordées, à l'échelle d'analyse et au contexte d'étude. Ces protocoles sont détaillés et justifiés dans le premier chapitre de ce travail, nous en dessinons ici succinctement les éléments principaux.

Le premier chapitre s'appuie principalement sur des données secondaires et constitue une revue de la littérature problématisée sur la sociogenèse de la méthode expérimentale par assignation aléatoire et *l'evidence-based policy*. L'abondance des travaux produits sur la méthode expérimentale par assignation aléatoire dans les pays anglo-saxons justifiait de ne pas y ajouter de nouveaux travaux empiriques, mais nécessitait une évocation des théories développées et des débats académiques suscités par cette méthode et son emploi dans le cadre de *l'evidence-based policy*. Nous avons procédé à la sélection de notre corpus par recherche bibliographique en croisant les bibliographies mobilisées par plusieurs articles clés jusqu'à arriver à saturation. Cette revue de la littérature s'est aussi enrichie des entretiens conduits avec nos interlocuteurs scientifiques ou politiques français⁴². Ces différents travaux sont étudiés et ordonnés selon nos intérêts de recherche, ce qui nous mène à affirmer, pour reprendre les termes d'H. Becker, que malgré les données secondaires mobilisées « *la table reste mienne même si quelques-uns de ses éléments sont préfabriqués* »⁴³.

Le second chapitre a utilisé l'entretien semi-directif à titre informatif et compréhensif, outil traditionnel de la sociologie de l'action publique, ainsi que l'analyse documentaire. Il s'est enrichi d'un recours à une cartographie des acteurs par les outils de l'analyse de réseau auxquels nous avons été formées dans le cadre de notre pratique évaluative. Nous avons dans un premier temps, procédé à une reconstruction factuelle de l'histoire du développement de cette méthode d'évaluation en France par un travail de triangulation entre les données issues d'entretiens et les documents puisés dans la littérature académique ou dans la littérature grise (documents produits par l'administration centrale). Dans un second temps nous avons procédé à une analyse compréhensive des entretiens afin de mettre à jour la diffusion de schéma cognitifs, de problématisations, et d'extraire les informations nous permettant de procéder à une cartographie des acteurs selon leurs liens institutionnels et interpersonnels.

Dans le dernier chapitre, nous avons eu recours à la méthodologie d'études de cas et à leur analyse croisée que nous avons apprise dans le cadre de notre pratique évaluative. Bien que la méthode d'étude de cas soit largement employée en science politique, comme le souligne J. Gerring, cette méthode fait encore peu l'objet de travaux méthodologiques précisant son opérationnalisation pratique⁴⁴. Dans un article consacré à

⁴² Une question introductive commune à tous nos entretiens était « À quel moment avez-vous entendu parler pour la première fois de la méthode expérimentale par assignation aléatoire et comment en êtes-vous venus à vous y intéresser ». Les réponses apportées permettaient de consolider notre historique de cette méthode ainsi que d'identifier les articles scientifiques clés.

⁴³ Howard Saul BECKER, *Écrire les sciences sociales*, Paris, Economica, 2004, p. 151.

⁴⁴ John GERRING, « What Is a Case Study and What Is It Good for? », (2004) 98-2 *The American Political Science Review* 341-354.

cette méthodologie, nous avons déjà évoqué comment les travaux méthodologiques en évaluation et en sociologie s'enrichissent mutuellement et apportent des orientations pour procéder à une analyse croisée de plusieurs unités de cas⁴⁵. Pour cette étude, nous avons sélectionné 15 cas d'expérimentations par assignation aléatoires conduites et menées à terme sur le territoire français et initiées entre 2006 et 2010. L'unité de chaque cas était l'expérimentation évaluée par assignation aléatoire en commençant l'analyse à partir de l'association entre acteurs de la mise en œuvre de l'intervention et évaluateurs, jusqu'à la production des résultats et leur prise en compte ou non par les acteurs de l'action publique. Ces études de cas ont été nourries par des entretiens auprès des parties prenantes des expérimentations (porteurs de projets, évaluateurs, financeurs, acteurs responsable de leur mise en œuvre), enrichis par une analyse documentaire des documents produits dans leur conduite (notes d'étape, rapports intermédiaires, rapports finaux et tout autre document de présentation ou d'organisation relatif au cas) ainsi que des articles scientifiques ou articles de presse les évoquant. Un protocole identique, systématique de collecte de données a été appliqué à chaque cas⁴⁶. Les différences de robustesse entre les cas proviennent du manque de disponibilité des acteurs ou de l'impossibilité de les identifier, ou du manque d'accès à la documentation liée au projet. Afin de tirer des enseignements généraux de ces cas, ils ont ensuite fait l'objet d'une analyse croisée par induction analytique afin de dégager des tendances générales ainsi que des typologies explicatives de certaines variations⁴⁷.

Cette approche empirique mobilise des méthodes qualitatives comprises au sens de A. Mucchielli comme *«une succession codifiée de processus de travail intellectuel proprement humain (comparaison, induction, généralisation, recherche de forme, invention de sens) (...) dans le but d'explicitier, en compréhension, à l'aide de concepts induits de l'observation, la structure intime et le fonctionnement interne d'un phénomène social»*⁴⁸. Le recours à des méthodes qualitatives s'explique par la nature de cet objet d'étude et sa problématique de recherche qui s'interroge sur les modalités d'avènement d'un phénomène ainsi que de ses effets sur l'organisation sociale. Une dimension quantitative existe dans cette recherche par l'étude des récurrences des arguments démonstratifs collectés afin d'établir leur fiabilité par la recherche d'une triangulation. Sa mobilisation, consiste ici à accorder plus de poids dans l'analyse aux éléments de preuve corroborés par plusieurs sources issues d'outils de collecte différents ou plusieurs sources issues des mêmes outils de collecte⁴⁹. Cependant, cette recherche demeure exclusivement qualitative si l'on se réfère à la définition instrumentale d'A. Mucchielli puisque le chercheur a été l'unique instrument de recueil et de traitement des données⁵⁰.

3.3. Une articulation du travail théorique et empirique par abduction

La présentation successive du cadre théorique puis des outils de la collecte de données de cette recherche peut donner l'apparence d'un travail hypothético-déductif, procédant

⁴⁵ Agathe DEVAUX-SPATARAKIS et Amandine GREGOT, « Les défis de l'emploi de l'étude de cas en évaluation », *Politiques Sociales et Familiales* 2012. p.33-43.

⁴⁶ Ce travail totalise au final 46 entretiens. Voir liste en annexe 2

⁴⁷ Cette démarche est explicitée en détail dans le chapitre 1.

⁴⁸ Alex MUCCHIELLI, *Les processus intellectuels fondamentaux sous-jacents aux techniques et méthodes qualitatives, Bilan et perspectives de la recherche qualitative en sciences humaines et sociales*, p. 23.

⁴⁹ On peut utiliser quatre types de triangulation : croiser plusieurs méthodes, plusieurs sources de nature différente issues de la même méthode, par exemple, plusieurs entretiens, plusieurs analystes pour corriger les biais d'interprétation, voire plusieurs théories explicatives. N. DENZIN, *Sociological Methods*, New York, McGraw-Hill, 1978; M. PATTON, *Qualitative Evaluation and Research Methods*, Thousand oaks, CA, Sage Publications, 2001.

⁵⁰ Alex MUCCHIELLI, *op. cit.* p. 23.

de manière linéaire et éprouvant la valeur explicative d'un modèle théorique aux données du terrain. Cependant, cette recherche a procédé à l'articulation du travail empirique et théorique par abduction à l'aide de méthode de collecte de données qualitatives. Ce manuscrit propose une présentation linéaire par souci de clarté afin d'en faciliter la lecture, mais ce travail de construction théorique et de collecte de données est le fruit d'un processus itératif d'influence de l'un sur l'autre étalé dans le temps. Lors de ce travail, il ne semble pas pertinent de distinguer une période de confrontation au terrain et de travail théorique correspondant soit à l'approche inductive soit à l'approche hypothético-déductive. La démarche abductive est apparue comme la plus en cohérence avec notre contexte de recherche ainsi que le choix du cadrage théorique. Comme l'a rappelé A. Labrousse dans le champ disciplinaire de l'économie, les théories institutionnalistes et leurs présupposés sur les comportements des acteurs, notamment celle de la rationalité limitée développée par Herbert Simon appellent une nouvelle logique d'investigation scientifique⁵¹. Cette démarche est en cohérence avec l'appareil méthodologique et théorique choisi puisqu'il correspond à sa conception constructiviste du rôle de l'acteur. En effet, il présuppose une prégnance du chercheur et son immersion dans un contexte et des construits cognitifs mais malgré tout son action de déduction, son « agence » pour reprendre les termes des théories néo-institutionnelles. Ceci ouvre alors sur une modification de la situation impliquant une circularité, une rétroaction sur les premières hypothèses.

Il serait trompeur de parler d'approche inductive puisque nous n'avons pas été en position phénoménologique "d'époque", c'est-à-dire capable d'aborder le terrain sans être influencé par notre connaissance des travaux sur notre objet d'étude. L'approche du terrain a été cadrée par des éléments théoriques notamment pour sélectionner les acteurs à interroger ainsi que structurer l'approche analytique de l'étude documentaire. Cette collecte de données était donc teintée de "préjugés" de la chercheuse dans sa grille de questionnement ainsi que dans sa grille d'analyse⁵². De plus, cette collecte de données, ayant eu lieu sur une période de trois ans, a contribué à constituer et choisir l'appareillage théorique jugé comme le plus pertinent pour aborder ce phénomène. Celui-ci a été modifié à plusieurs reprises afin de correspondre aux nouvelles données recueillies et analysées. Pour reprendre l'analyse de F. Guillemette, nous nous sommes placée dans une position d'induction si on la conçoit comme "*la détermination d'hypothèses compatibles avec les données échantillonnées*", mais dont "*le « test » constant de cette compatibilité est réalisé dans des « moments déductifs » qui constituent essentiellement une validation qui est proche de la vérification hypothético-déductive*"⁵³.

Il n'est donc pas pertinent non plus de parler d'approche hypothético-déductive puisque ce travail a cheminé par intuitions plus que par hypothèses, et ces intuitions ont été amendées au fur et à mesure de la confrontation aux données de terrain. L'inférence abductive a été mobilisée pour combiner les données empiriques et les cadres théoriques, afin de rester dans une attitude d'ouverture à la découverte de nouvelles compréhensions des phénomènes. Cette recherche qualitative s'est caractérisée par une souplesse de l'objet d'étude ajusté au vu des données de terrain et des avancements

⁵¹ Se référer directement à cet article libre d'accès sur le site de la revue, qui développe l'implication en science économique de la reconnaissance de l'autonomie des acteurs au-delà du modèle de l'*homo-economicus*, ainsi que l'intérêt des approches multi niveaux, et ce qu'elles suggèrent lorsqu'on les applique de manière réflexive au chercheur Agnès LABROUSSE, « Eléments pour un institutionnalisme méthodologique : autonomie, variation d'échelle, réflexivité et abduction », *Economie et Institutions* 2006.8.5-53, p. 38.

⁵² Marta ANADON et François GUILLEMETTE, « La recherche qualitative est-elle nécessairement inductive? », (2007) Hors Série-5 *Recherches Qualitatives* 26-37, p. 32.

⁵³ Marta ANADON et François GUILLEMETTE, *op. cit.* p. 33.

théoriques. Les hypothèses sont fondées empiriquement et le chercheur est dans un premier temps dans une logique exploratoire large, propice à l'émergence de surprises ou de faits étonnants. Le travail du scientifique consiste alors, à en formuler une théorie explicative qui est alors confrontée à nouveau aux données empiriques et permet ensuite l'émergence de nouvelles théories à éprouver. Ainsi, selon la perspective de Charles. S. Pierce l'abduction d'origine fait place à une logique déductive de test des hypothèses explicatives issues de l'abduction qui abouti enfin à une démarche inductive à la lumière des observations de terrain de l'émergence de nouvelles règles⁵⁴.

3.4. Un positionnement épistémologique de réalisme critique

En cohérence avec cette démarche itérative d'abduction, ce travail s'inscrit dans l'épistémologie du réalisme critique qui s'intéresse non pas à établir la généralisation d'une théorie mais à comprendre l'impact de mécanismes par leur confrontation à différents contextes.

L'épistémologie du réalisme critique comprend le « réel » comme un ensemble de structures et de mécanismes reliés de manière contingente à des événements empiriques⁵⁵. R. Bhaskar identifie trois niveaux de réalité, une première strate empirique, qui est celle observée par les individus ; une effective, qui désigne les événements se produisant ; et enfin le réel qui identifie les mécanismes générateurs de ces événements et les structures sur lesquels ils agissent⁵⁶. Cette épistémologie considère donc que le « réel » se compose de structures et de mécanismes générateurs détenant des caractéristiques intrinsèques, mais que ces dernières sont reliées de manière contingentes avec des événements empiriques⁵⁷. Les structures sont composées d'objets inter-reliés qui définissent leur identité les uns par rapport aux autres. Les mécanismes de structure génèrent des actions entre ces objets selon des conditions intrinsèques - quand il n'y a pas de changement de la nature du mécanisme - et des conditions extrinsèques - variant selon les contextes d'applications -⁵⁸. Ils peuvent intervenir à un niveau micro (individu), meso (gouvernance ou réseau) ou macro (institution plus large)⁵⁹. Ainsi, ce n'est pas parce qu'un événement généralement observé lorsqu'un mécanisme est en action est absent d'un cas, que la théorie de ce mécanisme est invalidée, l'absence de relation peut être expliquée par la contingence du contexte ou la présence de mécanismes compensateurs⁶⁰. L'étude de cas est donc particulièrement pertinente dans cette démarche épistémologique puisqu'elle permet d'observer le mécanisme dans une variété de contextes⁶¹. Cette démarche ne s'inscrit pas dans une conception mono-causale et linéaire de la causalité. Elle s'inscrit dans la notion de champ causal et théorise le lien non pas entre un effet et une cause unique, mais entre un effet et un faisceau causal réunissant un ensemble de causes, dont chaque cause ne peut

⁵⁴ Cette logique Piercéenne est développée dans l'article de Kruijff G.-J. M. (2005), « Peirce's late theory of abduction: A comprehensive account », *Semiotica*, 153-1/4, pp. 431-454 évoqué par Agnès LABROUSSE, « Eléments pour un institutionnalisme méthodologique : autonomie, variation d'échelle, réflexivité et abduction », *op. cit* p. 40.

⁵⁵ Eric W. K. TSANG et Kai-Man KWAN, « Replication and Theory Development in Organizational Science: A Critical Realist Perspective », (1999) 24-4 *The Academy of Management Review* 759-780.

⁵⁶ Roy BHASKAR, *A realist theory of science*, coll. Radical thinkers, n°29, London ; New York, Verso, 2008, p. 47.

⁵⁷ Marie-José AVENIER et Catherine THOMAS, « Mixer Quali et Quanti pour quoi faire? Méthodologie sans épistémologie n'est que ruine de la réflexion! », *Cahier de Recherche CERAG* 2011.

⁵⁸ Eric W. K. TSANG et Kai-Man KWAN, *op. cit*.

⁵⁹ A GEORGE et A BENNETT, *Case studies and theory development in the social sciences*, Cambridge, Mass, MIT Press, 2005, p. 147.

⁶⁰ Eric W. K. TSANG et Kai-Man KWAN, *op. cit*.

⁶¹ Marie-José AVENIER et Catherine THOMAS, *op. cit* p. 15.

à elle seule provoquer l'effet final, mais dont la synergie avec les autres causes est susceptible de créer l'effet en question⁶².

Cette épistémologie rejoint notre conception constructiviste de l'institution et de l'acteur puisqu'elle estime que bien que l'on puisse identifier des mécanismes communs sous jacent et apprendre sur leurs propriétés, pour autant ceux-ci ne sont activés que s'ils peuvent être appliqués dans des conditions et des structures leur permettant d'être activés. Il y a ainsi récursivité entre les mécanismes, la structure et le contexte. Elle nous permet donc de considérer qu'une connaissance proposant une montée en généralité peut être développée sur un mécanisme, et que celle-ci peut être notamment établie par l'étude de ce mécanisme dans plusieurs contextes. Cependant cela n'implique pas que ce mécanisme prenne la même forme dans la strate empirique observable par les individus. Notre travail a pour ambition, de mettre à jours les mécanismes à l'œuvre dans notre objet d'étude en s'intéressant particulièrement à sa diversité d'incarnation selon les acteurs promouvant son utilisation et selon les contextes et les structures sur lesquelles elle s'applique.

⁶² Barbara BEFANI, « Models of Causality and Causal Inference », dans *Broadening the Range of Designs and Methods for Impact Evaluation*, coll. DFID, 2012 p. 12.

***1^{ère} Partie : Une stratégie de recherche
élaborée par abduction: un choix
théorique articulé à la sociogenèse d'un
objet d'étude***

«La méthode des faits, pleine d'autorité et d'empire, s'arroge un air de divinité qui tyrannise notre créance, et impose à notre raison. Un homme qui raisonne, qui démontre même, me prend pour un homme: je raisonne avec lui; il me laisse la liberté du jugement; et ne me force que par ma propre raison. Celui qui crie voilà un fait, me prend pour un esclave.»

R. P. CASTEL, *L'Optique des couleurs, fondée sur les simples observations, et tournée surtout à la pratique de la Peinture, de la Teinture et des autres Arts coloristes*, Paris, 1740, p. 411

Considérer la politique du point de vue de la vérité revient à se placer en dehors du domaine politique. Ce point de vue est celui du diseur de vérité, qui trahit sa position – et ainsi la validité de ce qu'il a à dire- s'il tente d'interférer directement dans les affaires humaines et de parler le langage de la persuasion et de la violence.

Hannah ARENDT, Vérité et politique, dans *La Crise de la culture: huit exercices de pensée politique*, Paris, Gallimard, 1989 p.303.

“We must expand our students' vow of poverty to include not only the willingness to accept poverty of finances, but also a poverty of experimental results.”

Donald T Campbell et Julian C Stanley, *Experimental and Quasi-Experimental Design for Research*, Boston: Houghton Mifflin, 1963, p.3

L'introduction de ce travail a insisté sur l'importance que nous accordons à une conception constructiviste de l'acteur et de l'institution – au sens de la récursivité de l'influence de l'un sur l'autre – dans l'élaboration de notre stratégie de recherche, autant dans notre posture théorique qu'épistémologique ou méthodologique. Cette posture appelle à aborder notre objet d'étude par une approche abductive, c'est-à-dire à procéder par une démarche itérative entre premiers constats empiriques et élaboration du cadre théorique pertinent pour l'appréhender. L'étude de la diffusion de la méthode expérimentale en France et de ses effets, a pour avantage de pouvoir s'appuyer sur des phénomènes similaires dans d'autres contextes nationaux et internationaux. Cette méthode s'étant diffusée tardivement en France, un retour sur ses mécanismes de diffusions et son impact dans des contextes étrangers nous offre une fenêtre d'observation sur les mécanismes à l'œuvre dans ce phénomène. Elle ouvre des pistes de réflexions théoriques avant d'aborder le cas français et d'élaborer nos hypothèses de recherche. Nous proposons dans ce chapitre introductif de revenir dans un premier temps sur la sociogenèse à l'étranger de la méthode expérimentale par assignation aléatoire et de son utilisation dans le cadre de *l'evidence-based policy*. Les éléments de réflexion issus de ce travail nous permettent dans un deuxième temps, de détailler et de justifier notre élaboration théorique et méthodologique pour appréhender ce phénomène dans le contexte français.

Nous avons fait le choix dans notre introduction de ne pas procéder à un travail de définition de notre objet de recherche. Cette posture s'explique aussi par une conception constructiviste et interactionnelle de la définition des instruments d'action publique et des méthodologies. Nous estimons que la définition des objets étudiés a été évolutive, dans le temps, et qu'elle est issue d'un construit social produit par l'interaction des acteurs entre eux et une conjoncture particulière. Il nous apparaît donc indispensable de rendre compte de cette évolution et d'en expliquer les raisons afin d'appréhender ses derniers dans leur complexité. Ainsi, dans un premier temps, nous procédons à une étude de la sociogenèse de la méthode expérimentale ainsi que du courant de *l'evidence-based policy* montrant que ces deux objets sont intimement liés. En outre, cette démarche nous permet de comprendre la conjoncture internationale dans laquelle s'inscrit la diffusion en France de cette méthode comme instrument de *l'evidence based policy*.

A partir de ce premier développement, nous identifions les cadres théoriques et méthodologiques nous semblant les plus pertinents pour rendre compte de ce phénomène en France. Si, comme nous l'avons présenté en introduction, nous avons recours dans cette recherche de manière transversale aux concepts de *Champ d'action stratégique* et de *travail institutionnel*, ces derniers sont articulés et rendus opérationnels de manière différente dans nos deux chapitres empiriques. En effet, ces deux chapitres ont un objet d'intérêt et des questionnements différents, puisque le premier s'intéresse au phénomène d'émergence institutionnelle et le second à l'impact de cette construction institutionnelle. Pour chacune de ces démarches, nous articulons ces concepts avec les différents outils théoriques susceptibles d'être pertinents pour appréhender ces objets, en choisissant d'en écarter certains ou d'enrichir notre appareillage théorique avec d'autres approches. Enfin, nous présentons pour l'appareillage théorique de chaque chapitre les outils de collecte et d'analyse de données que nous avons jugé les plus opératoires et adaptés aux questions abordées dans nos hypothèses de recherche.

1. L'evidence-based policy et la méthode expérimentale par assignation aléatoire: sociogenèse de deux objets en interaction

À la fin de la décennie 2000, la méthode expérimentale par assignation aléatoire a été désignée par plusieurs instances de gouvernement nationales ou organisations internationales comme la méthode scientifique la plus rigoureuse pour établir l'efficacité d'une intervention publique mais aussi, la mieux à même de favoriser la prise en compte de travaux scientifiques par les acteurs de l'action publique⁶³. Cette méthode scientifique a été érigée comme l'*instrument* de choix de l'*evidence-based policy*, mouvement mettant l'apprentissage issu de travaux scientifiques au cœur de la fabrique des politiques publiques. L'ambition de ce chapitre est d'expliquer comment la méthode expérimentale par assignation aléatoire a été désignée par les acteurs de l'action publique comme la méthode scientifique la plus fiable et permettant de favoriser l'apprentissage politique à partir de travaux scientifiques.

Les *instruments de l'action publique* ont déjà été étudiés dans plusieurs travaux de sociologie de l'action publique proposant un programme de recherche autour de la déconstruction de ces objets⁶⁴. Cet agenda de recherche a pour but de dépasser la conception fonctionnaliste des instruments d'action publique qui considère ces dispositifs comme « neutres », exempts de théories et d'intérêts humains. Cette déconstruction consiste à mettre à jour les raisons poussant des instances de gouvernement à retenir des instruments plutôt que d'autres, à identifier les jeux sociaux, les acteurs de sa promotion ainsi que les cadrages cognitifs qui leur sont favorables et les théories qu'ils véhiculent⁶⁵. La méthode expérimentale par assignation aléatoire est un instrument particulièrement intéressant à déconstruire puisque cette technique, empruntée à la démonstration scientifique, est justement présentée comme un protocole garantissant une objectivité par le contrôle des différents biais de subjectivité. Ses promoteurs la défendent comme la méthode scientifique la plus neutre et donc la mieux à même de guider l'action publique sans être teintée d'idéologie ou sujette à des avis d'experts divergents. Utiliser le programme de recherche des instruments de l'action publique pour étudier cette méthode scientifique nous inscrit dans ce courant de recherche inspiré à l'origine par les travaux d'A. Desrosières sur la statistique et de M. Callon et B. Latour sur la construction sociale des expérimentations⁶⁶.

Nous souhaitons dans ce développement renseigner deux hypothèses issues du programme de recherche proposé par la sociologie de l'action publique s'intéressant aux instruments. La première hypothèse, est que :

⁶³ Peter ORSZAG, « Building Rigorous Evidence to Drive Policy | The White House »; INDEPENDANT EVALUATION GROUP, *World Bank Group Impact Evaluations: Relevance and Effectiveness*, Washington, D.C, World Bank, 2012.

⁶⁴ Patrick LE GALES et Pierre LASCOUMES, « L'action publique saisie par ses instruments », dans *Gouverner par les instruments*, Presse de Science po, 2004, p.13.

⁶⁵ *Ibid.*

⁶⁶ Alain DESROSIERES, « La statistique, outil de gouvernement et outil de preuve », dans *L'argument statistique*, Paris, Presses de l'école des mines, 2008; Michel CALLON, « Eléments pour une sociologie de la traduction, la domestication des coquilles Saint-Jacques et des marins-pêcheurs dans la baie de Saint-Brieuc », (1986) 36 *l'Année sociologique* pp 169-208; Madeleine AKRICH, Michel CALLON et Bruno LATOUR, « A quoi tient le succès des innovations? 1: L'art de l'intéressement », (1988) 11-12 *Gérer et comprendre, Annales des Mines*.

Hypothèse 1.1 : La méthode expérimentale par assignation aléatoire est un instrument porteur de théories sur, d'une part, les modalités de la démonstration scientifique et, d'autre part, sur l'action publique lorsqu'elle s'y applique.

Afin d'explorer cette hypothèse, nous nous intéressons à l'origine de cette méthode, les raisons de son émergence et les controverses scientifiques ayant jalonné son développement afin de mettre à jour ses caractéristiques. Nous étudions les acteurs ayant participé à son développement, à sa formalisation ainsi qu'à la définition de son périmètre d'utilisation et de pertinence.

Notre seconde hypothèse est la suivante :

Hypothèse 1.2 : L'alliance de cette méthode avec *l'evidence-based policy* résulte principalement d'un travail d'entrepreneurs scientifiques praticiens de cette méthode et bénéficiant d'un accès privilégié aux acteurs publics.

Nous nous intéressons ici aux processus de promotion et d'institutionnalisation de *l'evidence-based policy* par l'identification de ses acteurs et de leur provenance. L'étude de différentes incarnations de ce mouvement, nous permet d'identifier l'influence de ces groupes d'acteurs sur la mise en place par les instances de l'action publique d'une hiérarchie méthodologique en faveur de la méthode expérimentale par assignation aléatoire.

Bien que le travail de restitution que constitue la rédaction de thèse donne l'impression d'une conception stabilisée de l'objet de recherche, il semble ici important de souligner le processus d'élaboration de la progression de cette partie qui est révélatrice de la nature de nos résultats. Nous avons de prime abord, envisagé de présenter séparément les origines de la méthode aléatoire et celles de *l'evidence-based policy*, et de traiter ensuite conjointement leur rencontre. Lors de l'étude plus détaillée de la genèse de ces objets et notamment de leur environnement social, il est cependant apparu que la méthode expérimentale par assignation aléatoire et son statut de scientificité a justement émergé, dès son origine, par la rencontre de la science et du politique et que, réciproquement, le processus d'interaction entre la science et le politique ayant mené à *l'evidence-based policy* ne peut être compris sans un retour sur l'histoire de l'utilisation de l'expérimentation aléatoire et de sa promotion. Le choix a donc été fait de procéder par progression chronothématique en traitant l'émergence des deux objets en parallèle.

Dans une première partie, nous retraçons l'émergence de la méthode expérimentale par assignation aléatoire et son application à l'évaluation des interventions publiques. Dans un premier temps, en nous intéressant aux étapes clés du développement de la méthode expérimentale en science et ensuite de la formalisation par tirage aléatoire, nous soulignons la fonction de médiation de cette méthode d'une part entre le chercheur et son objet d'étude, et d'autre part entre le chercheur et la société. Cette méthode est porteuse d'ambivalences entre ses objectifs de production de connaissances et ceux de communication scientifique. Son développement en science l'inscrit dans une recherche du contrôle des biais de subjectivité liés à l'observation scientifique. L'expérimentation vise d'abord à contrôler les phénomènes observés afin d'en isoler les facteurs explicatifs pour que d'autres variables ne viennent perturber l'observation du scientifique. Elle organise ensuite, un protocole visant à réduire au maximum les interprétations ou

manipulations liées à la subjectivité du chercheur. Mais l'expérimentation est aussi un outil de communication scientifique qui a été historiquement développé dans une visée argumentative sur la démonstration de « l'évidence », c'est-à-dire sur l'administration d'une preuve scientifique pouvant être observée par des « profanes » - au sens de non scientifiques - et faisant justement appel à leur subjectivité, à leurs sensibilités, plus qu'à leur raison. Ce deuxième aspect de l'expérimentation inscrit son développement, dès ses origines, dans des logiques sociales et politiques qui contribuent à façonner les évolutions de cette méthode et sa position vis-à-vis des autres méthodes scientifiques. L'activité scientifique restant une activité sociale, la force de communication de cette méthode auprès des profanes contribue à établir la force de son administration de la preuve au sein de la communauté scientifique.

Dans un deuxième temps, nous nous intéressons aux modalités d'importations de cette méthode scientifique comme outil d'évaluation de l'action publique au cours du XXe siècle et particulièrement à partir des années 1960. Son transfert d'utilisation des sciences de la nature aux sciences humaines organise une redéfinition de son périmètre d'application pertinente. Son application aux sciences sociales requiert d'étudier les ajustements engendrés par cette importation et augmente l'influence du contexte social et politique sur sa validité scientifique. L'étude des différentes justifications successives de son utilisation, ainsi que des controverses qu'elle suscite, nous permet de mettre à jour les théories à la fois scientifiques, sociales et politiques véhiculées par cette méthode devenue instrument d'action publique.

Dans une deuxième partie, nous étudions la formalisation de la méthode expérimentale comme instrument de *evidence-based policy* à partir des années 2000. L'étude du cadrage cognitif opéré par les promoteurs de ce courant révèle la proximité de cette problématisation avec les théories portées par la méthode expérimentale par assignation aléatoire. Elle met à nouveau en lumière la dimension communicationnelle et sociale de cette méthode. Dans un premier temps, nous revenons sur la définition de ce mouvement de réforme de l'action publique et sur trois modèles de son incarnation en Grande-Bretagne, aux Etats-Unis et auprès des organisations internationales de l'aide au développement. Une approche interactionnelle de l'étude des acteurs et des enjeux rencontrés par les organisations ayant participé à la promotion de ce mouvement nous permet de comprendre la place privilégiée de la méthode expérimentale par assignation aléatoire dans ces différentes incarnations. Elle révèle aussi que ce mouvement, en plus de promouvoir un nouveau management de la connaissance issue de travaux scientifiques au sein des instances de l'action publique vise aussi, en retour, à exercer une influence sur les modalités de la production scientifique au sein de la communauté des chercheurs étudiant les interventions publiques, en les incitant à avoir recours à la méthode aléatoire.

Cette nouvelle ingérence politique dans les choix méthodologiques a été source d'un débat et de controverses parmi les chercheurs s'intéressant à l'action publique et/ou parmi les évaluateurs, sur lesquels nous revenons dans un deuxième temps. Ces débats précisent deux dimensions de l'expérimentation par assignation aléatoire. La première de nature épistémologique, concernant la conception de la causalité et de sa démonstration proposée par cette méthode, la seconde de nature sociale et politique défendant une certaine conception des objectifs de l'intervention publique et de son fonctionnement.

1.1. La méthode expérimentale par assignation aléatoire, un outil structurant de l'interaction entre la science et l'action publique

Ce premier développement propose un retour historique sur l'émergence de la méthode expérimentale par assignation aléatoire comme outil d'évaluation de politiques publiques. Cette recherche sur la genèse de cet instrument s'intéresse particulièrement aux différents discours et dynamiques qui ont accompagné la diffusion de cette méthode dans le champ scientifique, puis dans celui de l'évaluation des politiques publiques.

Cette méthode est d'abord restituée dans son contexte scientifique. L'expérimentation aléatoire s'inscrit dans la pratique scientifique expérimentale. Nous revenons brièvement sur les raisons d'émergence de cette pratique empirique en science et étudions l'argumentaire promouvant cette nouvelle démarche. Le développement de l'expérimentation est fortement lié à un processus de communication de la science : par la possibilité de l'observation sensorielle de la démonstration scientifique, elle en augmente la force de persuasion sur les publics non initiés. Elle porte aussi en elle la volonté de contrôler la subjectivité du chercheur qui se pose uniquement comme interprète du discours de la nature. Formalisée au début du XX^e siècle en agronomie par R. A. Fisher, la méthode expérimentale par assignation aléatoire d'entités constituant les groupes de comparaison, est destinée à contrôler davantage les biais d'interprétation, la subjectivité du chercheur et les variables externes pouvant interférer avec la démonstration.

Dès sa formalisation, cette méthode fut employée dans les sciences humaines afin notamment de contrôler les biais liés aux interprétations. Cependant, ces premières utilisations amenèrent à réviser les espérances de résultats produits par cette méthode, du fait de sa difficile mise en œuvre.

Sa conversion en outil d'évaluation des politiques publiques l'inscrit dans un contexte de forte nécessité de légitimation de nouveaux domaines d'intervention de l'État fédéral américain lié à une remise en cause de l'efficacité et de la pertinence de son intervention. Ses vertus scientifiques tenant d'une part au contrôle de la subjectivité des chercheurs dans la manipulation de son objet et l'analyse des résultats, et d'autre part la transparence revendiquée de son protocole, la positionnent en instrument argumentatif politique privilégié puisque garant de neutralité. Néanmoins, comme lors des premières expériences du début du siècle, cet emploi suscite des débats dans la communauté scientifique et politique, et ces réflexions sur les modalités de rencontres entre la science et la politique contribuent à faire émerger une théorisation de l'évaluation des politiques publiques dépassant l'application « brute » de méthodes scientifiques à l'action publique. La fin des années 1990, marque le retour de l'utilisation et de la promotion de cette méthode comme la plus rigoureuse dans la démonstration scientifique de la causalité. Sa théorisation est affinée par les économistes statisticiens qui en font leur outil de choix. Ses modalités d'interaction avec l'action publique sont redéfinies afin de garantir son intégrité scientifique.

1.1.1. Les origines d'un outil de production et de communication scientifique

La méthode expérimentale par assignation aléatoire s'inscrit dans l'héritage de la pratique expérimentale en science et a été théorisée comme une déclinaison de cette démarche permettant d'être au plus près des mêmes objectifs d'objectivité et de communication sociale des résultats. Un retour sur les raisons du développement de la pratique expérimentale scientifique permet d'identifier deux ambitions structurantes de cette dernière.

Tout d'abord, l'expérimentation témoigne d'une tension propre à l'entreprise scientifique entre le processus de production d'apprentissages scientifiques et leur communication à un plus large public pour la reconnaissance de la preuve administrée. Elle s'apparente à un outil de médiation, entre le chercheur et son objet d'une part, et le chercheur et la société d'autre part. Elle est présentée d'abord comme un outil de production de connaissance permettant de contrôler les biais liés à la subjectivité du chercheur, autant dans la manipulation, l'observation et l'interprétation des phénomènes naturels. Comme nous allons l'étudier ci-après, elle s'inscrit aussi, dès son origine, dans une stratégie de communication scientifique en proposant une incarnation de la démonstration scientifique observable par des publics scientifiques mais aussi, et surtout, non scientifiques, et se montre ainsi à même de les convaincre par un témoignage de l'évènement de l'administration de la preuve⁶⁷.

La pratique de l'expérimentation se caractérise, à ses débuts, davantage comme une démarche générale et multiple que par un protocole commun unifié. Cette diversité d'incarnations fragilise sa force argumentative et devient un obstacle à l'atteinte de ses objectifs puisqu'elle est source de désaccords sur le statut des démonstrations issues de pratiques variées, désaccords qui contribuent d'autant plus à brouiller sa force argumentative. La fin du XIX^e siècle, et surtout le début du XX^e siècle, sont des moments de théorisation des protocoles expérimentaux à même de garantir le meilleur contrôle de la subjectivité du chercheur, mais aussi de sortir l'expérimentation de l'environnement du laboratoire, tout en gardant le contrôle sur la manipulation du vivant et des variables externes. La méthode expérimentale par assignation aléatoire est alors théorisée par les travaux de R.A. Fisher comme la méthode à privilégier afin de garantir la plus rigoureuse application expérimentale sur le terrain (*field experiment*). Ce contact direct avec l'environnement extérieur, et notamment avec des sujets humains, place cette méthode en particulier au cœur des problématiques expérimentales de contrôle des biais subjectifs dans la production scientifique et de l'insertion de cette activité dans un contexte social, voire politique. L'objet de notre étude n'est pas de revenir de manière exhaustive et détaillée sur toutes les étapes du développement de cette pratique, mais de procéder à une analyse sélective privilégiant la mise en lumière de l'évolution de l'argumentaire scientifique justifiant sa supériorité épistémologique, ainsi que l'évolution des modalités de son utilisation politique.

1.1.1.1. L'expérimentation en science : un instrument d'accès contrôlé à la nature

Les premières traces de pratiques d'expérimentations ayant pour vocation d'apporter la preuve manifeste d'une démonstration scientifique sont présentes dès l'Antiquité,

⁶⁷ B LATOUR, « The force and the reason of experiment », dans Homer Eugene LE GRAND, *Experimental inquiries: historical, philosophical and social studies of experimentation in science*, Dordrecht; Boston; London, Netherlands ; Kluwer academic, 1990.

néanmoins ce n'est qu'à partir du XVII^e siècle, que cette pratique commence à être diffusée comme une méthode de démonstration scientifique supérieure aux autres modes de démonstration. Dès ses origines, elle est employée pour appuyer socialement l'administration d'une preuve scientifique. L'administration de la preuve est par nature une activité sociale, comme le soulignent M. Atten et D. Pestre « *la preuve est ce que quelqu'un administre, attendant des autres qu'ils jugent les choses de la même façon* »⁶⁸. Cette dimension sociale de la production de « faits » est aussi soulignée par H. Arendt dans son essai « Vérité et politique » elle explique : « *La vérité de fait, au contraire, est toujours relative à plusieurs : elle concerne des événements et des circonstances dans lesquels beaucoup sont engagés ; elle est établie par des témoins et repose sur des témoignages, elle existe seulement dans la mesure où on en parle, même si cela se passe en privé. Elle est politique par nature* »⁶⁹. L'expérimentation, par sa capacité à incarner la démonstration, s'illustre par son efficacité à convaincre ses témoins. Contrairement à l'argumentation philosophique dont la démonstration fait appel au cheminement de la Raison, l'expérimentation quant à elle mobilise l'observation sensorielle, au témoignage de l'auditoire. B. Latour souligne que dès son origine, l'expérimentation, notamment celle d'Archimède, met en scène des démonstrations spectaculaires afin de convaincre le pouvoir de la véracité de sa théorie du levier⁷⁰. Ce retour historique amène cet auteur à estimer que « *le montage expérimental tout entier est depuis l'origine un stratagème pour marquer un point dans le combat qui mène au convaincre* »⁷¹.

L'étude de S. Shapin et S. Schaffer sur la controverse qui, dans la deuxième partie du XVII^e siècle, opposa Thomas Hobbes, défenseur de la raison philosophique, et Robert Boyle, promoteur de la démarche expérimentale, témoigne de l'argumentaire accompagnant la légitimation de l'expérimentation⁷². Cette controverse est particulièrement intéressante pour notre étude puisque c'est à la suite de celle-ci que l'expérimentation sera reconnue comme une méthode scientifique supérieure à la démonstration analytique philosophique. L'expérimentation est employée dans un double but, l'un d'apporter une démonstration de la causalité par la présentation d'une preuve dépassant la subjectivité du chercheur et, l'autre d'incarner et de matérialiser la démonstration scientifique afin d'en augmenter l'audience et donc l'impact sur la société. L'expérimentation devient une ressource puissante de persuasion qui s'inscrit dans le mythe des capacités du scientifique à « lire le livre de la Nature » grâce aux techniques de l'expérimentation⁷³. Dans ses écrits, R. Boyle formule cette ambition : « *je m'estimerai tout à fait satisfait si l'on pensait que je n'ai guère regardé d'autres livres que celui de la Nature* ».

T. Hobbes critique la méthode expérimentale qu'il assimile à une expérience individuelle et donc soumise aux apparences et illusions⁷⁴. Selon l'auteur du Léviathan, le recours à la raison, faculté partagée par tous les hommes, est le seul moyen permettant

⁶⁸ Michel ATTEN et Dominique PESTRE, *Heinrich Hertz: l'administration de la preuve*, Paris, Presses Universitaires de France, 2002, p. 13.

⁶⁹ Hannah ARENDT, « Vérité et politique », dans *La Crise de la culture: huit exercices de pensée politique*, Paris, Gallimard, 1989 à la page 303.

⁷⁰ Dans l'introduction de ce chapitre, l'auteur explique comment Archimède a démontré le pouvoir du levier en tirant par la force de poulies un navire marchant. B LATOUR, « The force and the reason of experiment », *op. cit.* p.49

⁷¹ Citation dans l'ouvrage de Jean-Michel BERTHELOT, *L'emprise du vrai: connaissance scientifique et modernité*, Paris, Presses Universitaires de France, 2008, p. 58.

⁷² Steven SHAPIN et Simon SCHAFFER, *Leviathan and the air-pump: Hobbes, Boyle, and the experimental life*, Princeton, NJ, Princeton University Press, 1985.

⁷³ David GOODING et Simon SCHAFFER, *The Uses of experiment: studies in the natural sciences*, Cambridge [England]; New York, Cambridge University Press, 1989, p. 6.

⁷⁴ Steven SHAPIN et Simon SCHAFFER, *op. cit.* p. 103.

d'atteindre un consensus et donc la paix dans le contexte de la fin de la guerre civile qui déchira la Grande-Bretagne⁷⁵. Aussi, il s'oppose politiquement à concéder une place aux institutions religieuses dans la gestion de l'État, celle-ci devant être réservée au souverain. À l'inverse, R. Boyle estime que l'expérimentation permettra de clore les querelles dans la communauté scientifique et donc dans la société. Sous la période de restauration durant laquelle l'Église anglicane occupe le pouvoir, la démonstration de l'existence du vide par la pompe à air, proposée dans les expérimentations de R. Boyle, conforte cette position de l'Église, par l'identification d'un espace non matériel, et donc intemporel, justifiant le pouvoir spirituel de l'Église. La pratique de la « philosophie » expérimentale permet ainsi à Boyle à la fois de prendre ses distances avec la philosophie marquée par T. Hobbes, qui défend une conception moniste soutenant que l'âme n'existe pas séparément du corps ; et en outre, de légitimer la place de l'Église⁷⁶. D'ailleurs Boyle s'approprie le vocabulaire ecclésiastique et présente les expérimentalistes comme les « *prêtres de la nature* »⁷⁷.

Selon les auteurs de cet ouvrage, la sortie victorieuse de la méthode expérimentale à l'issue de cette controverse tient donc davantage au soutien politique environnant les travaux de R. Boyle, qu'à ses qualités intrinsèques. Ils identifient trois technologies propres à l'expérimentation. Une technologie littéraire, qui consiste à restituer les expérimentations de manière très détaillée et réaliste afin que les lecteurs puissent avoir l'impression de l'avoir vécu. Une technologie sociale, où interviennent des témoins de qualité, des *gentlemen*, afin d'augmenter la légitimité de la démonstration. Enfin, une technologie matérielle, puisque les expérimentations ont recours à la construction de nouveaux appareils techniques comme media entre l'homme et la nature. Cet appareillage est essentiel puisque la solidité du fait tient à l'absence d'intervention humaine dans leur genèse⁷⁸. Cet appareillage coûteux a été une des raisons de la création des sociétés scientifiques à cette période et notamment de la Royal Society créée au début des années 1660 pour financer l'entreprise expérimentale⁷⁹. Ce recours aux appareils a d'ailleurs été au cœur des critiques de T. Hobbes qui estime qu'ils contribuent à créer artificiellement des phénomènes n'existant pas dans la nature.

Cette controverse d'abord scientifique, devient donc sociale, puis politique, c'est l'organisation de la société qui est en jeu ainsi que le statut de l'Église. L'expérimentation s'inscrit alors dans un débat d'argument mobilisant la société, et donc des stratégies communicationnelles pour convaincre le public mondain. Si l'expérimentation prend place d'abord en laboratoire et peut être sujette à l'échec et aux tâtonnements, c'est seulement une fois stabilisée qu'elle est rendue publique afin de procéder à la démonstration scientifique⁸⁰. L'expérimentation a donc un double emploi, l'élaboration de connaissance et la communication des résultats. Cette fonction communicative de l'expérimentation est critiquée fortement par G. Bachelard, qui évoque ces pratiques expérimentales publiques qui se sont prolongées au XVIII^e siècle en ces termes « *en donnant une satisfaction immédiate à la curiosité, en multipliant les occasions de la curiosité, loin de favoriser la culture scientifique on l'entrave. On remplace la connaissance par l'admiration, les idées par les images.* »⁸¹

⁷⁵ Steven SHAPIN et Simon SCHAFFER, *op. cit.* p. 155.

⁷⁶ Steven SHAPIN et Simon SCHAFFER, *op. cit.* p. 294

⁷⁷ *Ibid.* p. 319

⁷⁸ *Ibid.* p. 30

⁷⁹ *Ibid.* p. 42.

⁸⁰ Steven SHAPIN, « The House of Experiment in Seventeenth-Century England », (1988) 79-3 *Isis* 373- 404, 400.

⁸¹ Gaston BACHELARD, *La formation de l'esprit scientifique: contribution à une psychanalyse de la connaissance*, Paris, J. Vrin, 1934, p. 29.

Au cours du XVIII^e et XIX^e siècle, l'expérimentation est formalisée comme une méthode de démonstration et uniformisée par la définition du protocole expérimental, notamment par les travaux de Claude Bernard en médecine. Elle est structurée autour d'une pratique commune suivant trois éléments selon B. Latour. En premier lieu, l'organisation d'un test, puis un rapport écrit de ce test et de son protocole, et enfin la confortation ou l'infirmité d'un point de la littérature scientifique par cette démonstration⁸². Ce dernier identifie trois moments de l'expérimentation, sa mise en place par le scientifique ; la réaction de l'objet à cette expérimentation ; et la communication de ses résultats et sa reconnaissance par les pairs. La force persuasive de l'expérimentation réside à l'extérieur de sa conduite, elle est dans le jugement des autres scientifiques sur la qualité de l'expérimentation⁸³. Ce jugement s'appuie sur des critères tenants au contrôle des biais liés à l'environnement, au contrôle du phénomène artificiel et enfin à la subjectivité du chercheur sur l'interprétation des résultats, lequel doit procéder « les yeux bandés »⁸⁴.

En 1843, dans son ouvrage « System of logic », John S. Mill établit la supériorité de l'expérimentation sur l'observation comme universellement reconnue et présente plusieurs méthodes de recherche expérimentales⁸⁵. L'expérimentation est présentée comme la méthode la plus simple pour démontrer une causalité par sa capacité à renverser le procédé et produire un effet artificiellement⁸⁶. La maîtrise complète par le chercheur des éléments mobilisés pour créer artificiellement le phénomène est présentée comme fondamentale, afin de s'assurer que l'effet produit provient bien de l'antécédent (ce qui enclenche la cause) et non pas aux moyens artificiels employés pour créer l'antécédent⁸⁷. Néanmoins, ce philosophe garde toujours une place, dans la démonstration de la causalité, à l'observation, bien que celle-ci dépende des diversités de configurations présentes dans la nature et soit donc plus difficile à rencontrer⁸⁸.

Ainsi, à la lumière de cette littérature, le développement de la méthode expérimentale en science s'inscrit dans deux logiques paradoxales. L'une est liée à la recherche de la causalité par la réplication artificielle d'un phénomène. La pratique expérimentale vise alors à contrôler les biais liés à l'interférence de la subjectivité du chercheur, que ce soit dans la manipulation expérimentale ou dans l'interprétation des résultats, ainsi qu'à contrôler les variables extérieures pouvant exercer une influence sur le phénomène observé. L'autre est liée à la nécessité de rendre la démonstration scientifique accessible à un plus large public afin d'augmenter la force de la preuve administrée. La pratique expérimentale tend alors à inscrire sa démonstration dans des contextes politiques et

⁸² Bruno LATOUR, « The force and the reason of experiment », *op. cit.* p. 57.

⁸³ Harry COLLINS, *Changing order: Replication and Induction in Scientific Practice*, University Of Chicago Press, 1985 (p83-84) cité par David GOODING et Simon SCHAFER, *The Uses of experiment: studies in the natural sciences*, Cambridge [England]; New York, Cambridge University Press, 1989, p. 12.

⁸⁴ Ian HACKING, « Telepathy: Origins of Randomization in Experimental Design », (1988) 79-3 *Isis* 427-451, p.428.

⁸⁵ Voir le Livre III de l'induction, chapitre sept « de l'observation et de l'expérimentation »

⁸⁶ « Supposons que, par la comparaison des cas de l'effet, on ait trouvé un antécédent qui y paraît être, et y est peut-être, invariablement lié, on n'a pas encore par là la preuve que cet antécédent est la cause, tant qu'on n'a pas renversé le procédé et produit l'effet par le moyen de l'antécédent. Si l'on peut produire l'effet artificiellement, et si alors l'effet s'ensuit, l'induction est complète ; cet antécédent est la cause de ce conséquent ⁸⁶. Mais, ici, on a ajouté le témoignage de l'expérimentation à celui de la simple, observation », Livre III de l'induction, chapitre sept « de l'observation et de l'expérimentation » paragraphe 4.

⁸⁷ En effet J. S. Mill rajoute à la citation précédente en note de bas de page « A moins que le conséquent ne provint pas de l'antécédent même, mais des moyens employés pour créer l'antécédent. Comme, cependant, ces moyens sont en notre pouvoir, il y a une probabilité que nous en ayons aussi une connaissance suffisante pour nous faire juger si c'est là le cas ou non. »

⁸⁸ « S'il arrive que la nature nous offre des cas suffisamment variés dans leurs circonstances, et si nous réussissons à découvrir, soit parmi les antécédents prochains, soit dans quelque autre ordre d'antécédents, quelque chose qui se rencontre toujours quand l'effet a lieu, quelques différentes que soient les circonstances, et qui ne se rencontre jamais quand l'effet n'a pas lieu, nous pouvons alors découvrir par la simple observation, et sans expérimentation, une loi réelle de la nature. » Livre III de l'induction, chapitre sept « de l'observation et de l'expérimentation » paragraphe 4.

sociaux et à faire appel dans sa légitimation à la subjectivité de témoins. Son accès, supposé « direct » aux lois de la Nature, lui donne une force argumentative sans précédent pour clore les controverses scientifiques. La méthode expérimentale par assignation aléatoire s'inscrit dans cette double logique. Son protocole est présenté par ses promoteurs comme le mieux à même d'atteindre l'objectif de contrôle de la subjectivité du chercheur et sa reconnaissance, comme méthode reine de l'expérimentation, est inscrite dans une dynamique d'association dès ses origines avec les pouvoirs publics.

1.1.1.2. La méthode expérimentale par assignation aléatoire, comme protocole expérimental par excellence

La méthode expérimentale par assignation aléatoire propose un protocole permettant une nouvelle correction des biais subjectifs. Elle permet aussi de sortir l'expérimentation du laboratoire, non pas uniquement pour rendre publique la démonstration comme aux siècles précédents, mais pour augmenter la force de démonstration de cette dernière en procédant à une production expérimentale sur le terrain. La théorisation de la méthode expérimentale par assignation aléatoire est souvent attribuée aux travaux de l'agronome R. A. Fisher dans les années 1930. Bien que ce dernier en ait explicité les bases et justifications statistiques, le tirage aléatoire dans les expérimentations hors du cadre du laboratoire s'était déjà développé à partir de la deuxième moitié du XIX^e siècle dans les sciences humaines. Cette méthode ne s'est donc pas imposée radicalement comme la norme de la rigueur expérimentale, mais elle résulte d'un travail de tâtonnement dans les pratiques expérimentales sur une période plus longue.

En effet, dans un premier temps se développe l'expérimentation sur le terrain par comparaison d'un groupe témoin et d'un groupe test. Ces premières utilisations de cette démarche de comparaison sont identifiées dès le XVIII^e siècle par l'expérimentation de James Lind en 1747 qui a démontré l'efficacité des agrumes sur des marins atteints de Scorbut⁸⁹. Cette pratique se diffuse au XIX^e siècle en agronomie, puis en biologie avec notamment l'expérimentation de Louis Pasteur, en 1882, destinée à tester l'effet d'un vaccin en l'allouant à un groupe de moutons et non à un autre. Dans cette entreprise, l'expérimentation est mobilisée pour convaincre l'opinion de l'efficacité du vaccin ainsi que pour répondre au scepticisme d'autres chercheurs. Son déroulement fait l'objet de rapports quotidiens dans les journaux français et anglais, et conclut à un succès de L. Pasteur⁹⁰.

Ensuite, dans le domaine de la psychophysique, le tirage au sort s'est développé dans la deuxième partie du XIX^e siècle, dans la pratique expérimentale de laboratoire, afin de contrôler les biais liés aux sujets participants⁹¹. Charles S. Peirce et Joseph Jastrow, dès les années 1880, lors d'une expérimentation sur la capacité à ressentir des différences de poids entre différents objets, tirent au sort l'ordre de présentation des objets et s'appuyant sur la loi des erreurs, établissent que leur expérience démontre un continuum de baisse de capacités de prédiction de l'estimation⁹². Le tirage aléatoire se diffuse ensuite dans la recherche psychique, organisant des expérimentations pour identifier

⁸⁹ Thomas DUNCAN, « Sailors, Scurvy and Science », *Journal of the Royal Society of Medicine* 1997.90.

⁹⁰ Steven LEVITT et John LIST, *Field Experiments in Economics: the Past, the Present, and the Future*, National Bureau of Economic Research, Cambridge MA, 2008, p. 6.

⁹¹ La psychophysique est une branche de la psychologie expérimentale qui identifie les relations quantitatives entre un stimulus physique et la perception que l'on en a.

⁹² Cette méthode fût ensuite mobilisée pour conduire des expérimentations permettant de déterminer les pouvoirs télépathiques Ian HACKING, « Telepathy: Origins of Randomization in Experimental Design », *op.cit.* p. 432.

l'absence ou la présence de capacités télépathiques⁹³. Peu à peu, le recours à l'expérimentation et au tirage aléatoire, pour contrôler des biais d'observation, se diffuse auprès des chercheurs en psychologie jusqu'au début du XX^e siècle⁹⁴. Selon les travaux de plusieurs historiens des sciences, il semblerait donc que ce soit plutôt R. A. Fisher qui se soit inspiré des travaux expérimentaux ayant recours au tirage aléatoire dans le champ de la psychologie, que l'inverse⁹⁵.

R. A. Fisher reste malgré tout le théoricien fondamental de la méthode expérimentale par assignation aléatoire. Dans son ouvrage « the design of experiments », il en définit le protocole, le périmètre de validité et le type de questions auxquelles elle est capable de répondre. Son ambition est de proposer une méthode permettant de clore les débats de l'époque sur les résultats d'expérimentations dont soit l'interprétation, soit le protocole sont critiqués⁹⁶. Il souhaite stabiliser un protocole expérimental permettant de réduire l'interprétation du chercheur et les hypothèses de recherche, et d'autre part d'estimer mathématiquement la marge d'erreur du résultat produit. De plus, cette méthode est présentée comme économique, puisqu'elle permet d'améliorer l'efficacité démonstrative d'une expérimentation et nécessite moins de répliques pour arriver à un niveau de significativité. Suivant la théorie des erreurs de Gauss, le tirage au sort permet à la fois de réduire les effets de facteurs autres que celui observé par l'expérimentation en moyennant les différences et permet en outre de procéder à un calcul de probabilité de marge d'erreur. Dans cet ouvrage, R. Fisher redéfinit ce qu'il nomme « *the art of experimental design* » de la manière suivante : « *les observations expérimentales sont uniquement des expériences soigneusement planifiées en amont et conçues pour constituer une base solide de nouvelle connaissance; c'est-à-dire qu'elles sont systématiquement liées à un corpus de connaissances déjà établies, et que les résultats sont intentionnellement observés et enregistrés avec précision* »⁹⁷. Cette définition reprend donc les caractéristiques de l'expérimentation identifiées par B. Latour, à savoir, l'ancrage dans la littérature scientifique, l'observation empirique et le récit détaillé du déroulement de l'expérimentation ainsi que du protocole mis en place. Pour expliquer sa démonstration, R.A. Fisher mobilise une anecdote pour incarner ce protocole : « *the lady tasting tea* » (la dame goûtant du thé). Un retour sur cet exemple permet au lecteur non spécialiste des méthodes statistiques de comprendre cette démarche.

Encadré 1 : "The lady tasting tea" la logique expérimentale proposée par R. A. Fisher

The lady tasting tea

Afin d'expliquer la démarche de la méthode expérimentale par assignation aléatoire, A. Fisher restitue une expérience informelle conduite dans le cadre de ses recherches. Le contexte expérimental est le suivant : une des collègues d'R. A. Fisher est offensée qu'on lui ait servi sa tasse de thé au lait en versant d'abord le lait et le thé ensuite. Cet ordre semble anodin pour un lecteur français, mais un lecteur familier de la culture britannique reconnaîtra que rien n'est laissé au hasard dans ce rituel culturel. Il est en effet convenu outre-Manche que les personnes appartenant aux classes supérieures boivent ce breuvage en ajoutant le lait une fois le thé servi et non l'inverse. Devant les réprobations de ces collègues, cette « lady » avance qu'elle peut reconnaître au goût la différence entre les deux manières de

⁹³ Elles procédaient par tirage aléatoire de cartes dont il fallait deviner la couleur. Pour avoir un récit en détail de ces expérimentations voir Ian HACKING, « Telepathy: Origins of Randomization in Experimental Design », *op.cit.*

⁹⁴ Trudy DEHUE, « Deception, Efficiency, and Random Groups: Psychology and the Gradual Origination of the Random Group Design », (1997) 88-4 *Isis* 653- 673.

⁹⁵ *Ibid.*; Ian HACKING, « Telepathy: Origins of Randomization in Experimental Design », *op.cit.*

⁹⁶ Ronald A. FISHER, *The Design of Experiments*, New York, Hafner Publishing Company, 1935, p. 1.

⁹⁷ Soulignage ajouté par l'auteure du présent texte, traduction à partir de l'anglais: « *experimental observations are only experience carefully planned in advance and designed to form a secure basis of new knowledge ; that is, they are systematically related to the body of knowledge already acquired, and the results are deliberately observed, and put on record accurately* » *Ibid.*, p. 8.

servir. Ce pari est l'occasion pour R.A. Fisher de mettre en application son protocole expérimental aléatoire. On peut noter ici l'affiliation de cette expérience avec le développement de cette méthode dans le champ de la psychophysique exposé précédemment, ainsi que l'importance du statut qualifié des témoins de l'expérimentation présente dès le XVIII^e siècle dans cette pratique⁹⁸.

R. A. Fisher propose le dispositif suivant. Il prépare huit tasses dont quatre de chaque sorte, qui sont proposées à sa collègue de manière aléatoire. Celle-ci après les avoir goûtées doit alors les répartir dans deux groupes de quatre tasses. Fisher compte 70 façons possibles de choisir les tasses et explique que plus le nombre de tasses est élevé moins l'identification de l'ordre de la composition du breuvage peut être dû à la chance. En ayant recours aux probabilités et particulièrement à la loi normale de Gauss, Fisher établit que sa collègue prouvera ses capacités gustatives que si elle reconnaît parfaitement les huit tasses. En appliquant la marge conventionnelle des 5% de marge d'erreur, il estime que la reconnaissance de trois tasses sur quatre n'est pas significative puisqu'elle a 16 chances sur 70 d'être rencontrée (soit plus de 20%).

Tasses identifiées avec succès	Configurations Possibles ✓ succès x échec	probabilités
0	x x x x	1x1=1
1	x x x ✓ / x x ✓ x / x ✓ x x / ✓ x x x	4x4=16
2	✓ ✓ x x / ✓ x ✓ x / ✓ x x ✓ / x ✓ x ✓ / x x ✓ ✓ / x ✓ ✓ x	6x6=36
3	x ✓ ✓ ✓ / ✓ x ✓ ✓ / ✓ ✓ x ✓ / ✓ ✓ ✓ x	4x4=16
4	✓ ✓ ✓ ✓	1x1=1
	total	70

Il choisit de formuler une hypothèse nulle (au sens de négative). Si l'on formulait une hypothèse positive « la dame ne se trompe jamais pour déterminer l'ordre de la composition du breuvage » une seule erreur invaliderait l'hypothèse et elle ne pourrait être prouvée par des expériences à répétition. L'hypothèse nulle qu'il formule est donc « la dame n'est pas capable de déterminer l'ordre de la composition du breuvage » et alors une expérience pourra apporter des premiers éléments de réponse concluants.

Il précise alors qu'il faudrait dans l'idéal que toutes les tasses soient identiques en tous points sauf sur l'élément testé, ce qui n'est pas possible. Il choisit alors les tasses de manière aléatoire avant de les remplir. Le nombre de 4 et 4 étant fixé, on entend par là que lorsque l'on cherche à composer le premier mélange « thé avant lait » on tire au sort parmi les huit tasses et ainsi de suite. Fisher précise que ce tirage au sort doit être le dernier phénomène dans « l'histoire physique des objets » avant qu'ils soient soumis au test afin de ne pas affecter la réaction expérimentale⁹⁹.

Le résultat de l'expérimentation n'est pas présenté dans cet ouvrage, mais selon un des témoins, leur collègue a réussi à démontrer avec succès son argument¹⁰⁰. Fisher précise néanmoins, qu'il serait souhaitable d'augmenter la taille de l'échantillon afin de rendre l'expérimentation plus sensible ou de réitérer l'expérimentation afin d'agrèger les résultats ou qu'il n'y ait pas de nombre préétabli de tasses de chaque sorte ce qui réduirait les chances de tomber juste à 1 sur 256.

Cependant, à l'époque, cette démonstration ne reçoit pas un accueil scientifique favorable auprès des statisticiens britanniques qui lui préfèrent les méthodes statistiques d'appariement quasi expérimentales¹⁰¹. C'est par son utilisation dans les sciences humaines, et particulièrement celles associées à la décision publique, que cette méthode se développe dans la première moitié du XX^e siècle.

⁹⁸ R.A. Fisher y fait lui-même référence « returning to the possible results of the psychophysical experiment » Ronald A. FISHER, *The Design of Experiments*, op. cit. p. 11.

⁹⁹ Ibid. p. 20.

¹⁰⁰ David SALSBURG, *The lady tasting tea: how statistics revolutionized science in the twentieth century*, New York, W.H. Freeman, 2001.

¹⁰¹ Dans son article I. Hacking précise qu'un des étudiants de Fisher souhaitant défendre une thèse sur la méthode aléatoire en 1932 n'a pu trouver aucun chercheur souhaitant participer au jury Ian HACKING, « Telepathy: Origins of Randomization in Experimental Design », op. cit. p. 429.

1.1.1.3. Une formalisation du tirage aléatoire par son application à l'action publique

Au début du XX^e siècle, à partir de la présidence d' H. Hoover et particulièrement sous la présidence de F. D. Roosevelt, un contexte américain favorable à une intervention du gouvernement fédéral dans l'organisation du « bien-être » (*welfare*) de sa population et le développement des sciences humaines et sociales, favorise l'émergence de mesures administratives ayant vocation à s'appuyer sur des faits produits par des méthodes scientifiques¹⁰². À la fin des années 1930, 8 000 chercheurs en sciences sociales travaillent pour le gouvernement fédéral et ce chiffre double pendant la Seconde Guerre mondiale¹⁰³. Ce développement est particulièrement présent en éducation, domaine investi par les psychologues du fait de la proximité entre l'étude des phénomènes mentaux et celle des capacités mentales ayant vocation à être augmentées par l'enseignement public. À cette occasion, les psychologues sortent les expérimentations de leur laboratoire pour investir les salles de classe. La question du contrôle des variables extérieures devient alors centrale. Ces derniers importent dans les sphères de gouvernement leurs outils et contribuent à promouvoir la méthode aléatoire comme la méthode idéale en termes de rigueur et de contrôle des biais¹⁰⁴. C'est à ce moment que W.A. McCall introduit, en 1923, la composition des groupes par tirage aléatoire qu'il présente comme une solution à même de contrôler ces biais et plus économique que l'appariement, fondé sur la base de pré-test auprès des élèves permettant de déterminer des profils comparables¹⁰⁵. Le développement de méthodes quantitatives en psychologie ainsi que la recherche de contrôle de l'expérimentation, s'inscrit dans cette nouvelle demande politique d'appuyer sa légitimité sur des données impartiales¹⁰⁶. C'est la rencontre de l'expérimentation psychologique et de la bureaucratie managériale en éducation qui favorise la formalisation et le développement de la méthode expérimentale par assignation aléatoire dans ce domaine. W.A. McCall, dans son ouvrage « *experiment in education* » défend le tirage aléatoire comme méthode permettant, d'une part de faire des économies en action publique en choisissant ce qui est efficace, et d'autre part comme un outil pour annuler les biais liés à la subjectivité à la fois des chercheurs et des sujets y participant¹⁰⁷. Ces travaux provoquent un engouement pour cette méthode dans le domaine de l'éducation, mais font rapidement place à un rejet de cette dernière dès 1935, alors que ses défenseurs les plus convaincus reconnaissaient que « *les contributions directes venant des expérimentations contrôlées ont été décevantes* »¹⁰⁸.

¹⁰² Cette association s'incarne dans la collaboration avec notamment le chercheur C. E. Merriam membre du Research Committee on Social Trends sous Hoover et du National Resources Planning Board sous F. D. Roosevelt. Ce chercheur en science politique nourrissait un intérêt pour les méthodes de recherches en psychologie et est identifié comme le « père » du mouvement comportementaliste en science politique. Sa proximité avec Roosevelt était telle que ce dernier l'appelait « uncle Charley » Trudy DEHUE, « Deception, Efficiency, and Random Groups: Psychology and the Gradual Origination of the Random Group Design », (1997) 88-4 *Isis* 653- 673, p.662.

¹⁰³ Trudy DEHUE, « Establishing the experimenting society: The historical origin of social experimentation according to the randomized controlled design », (2001) 114-2 *American Journal of Psychology* 283- 302, p.293.

¹⁰⁴ Trudy DEHUE, « Deception, Efficiency, and Random Groups: Psychology and the Gradual Origination of the Random Group Design » *op. cit.* p.672.

¹⁰⁵ Trudy DEHUE, « Deception, Efficiency, and Random Groups: Psychology and the Gradual Origination of the Random Group Design » *op. cit.* p. 668

¹⁰⁶ *Ibid* p. 671

¹⁰⁷ Trudy DEHUE, « Establishing the experimenting society: The historical origin of social experimentation according to the randomized controlled design », (2001) 114-2 *American Journal of Psychology* 283- 302, p. 291.

¹⁰⁸ Donald T CAMPBELL et Julian C STANLEY, *Experimental and Quasi-Experimental Design for Research*, Boston, Houghton Mifflin, 1963, p. 2.

En revenant sur cet épisode dans leur ouvrage de 1963, D.T. Campbell et J. Stanley tirent les enseignements pour de nouvelles applications. Ils estiment que cet échec tient aux trop fortes attentes envers les expérimentations et à des problèmes de rigueur de conception des protocoles. La première erreur est imputée aux promoteurs de cette méthode qui ont trop fortement déprécié les résultats scientifiques venant d'autres méthodes de production de connaissances et émettent l'hypothèse que l'absence d'amélioration des techniques d'enseignement tient uniquement au manque d'application des méthodes scientifiques adéquates dans ce domaine¹⁰⁹. Campbell et Stanley proposent de revoir à la baisse les attentes envers la production expérimentale en concluant : « *nous devons étendre le vœu de pauvreté de nos étudiants pour inclure non seulement la volonté d'accepter une pauvreté financière, mais aussi une pauvreté des résultats expérimentaux* »¹¹⁰. Ils appellent à une multiplication des expérimentations afin de consolider les résultats et déconseillent l'emploi de cette méthode pour trancher entre deux points de vue opposés sur une même intervention, puisque chaque point de vue est susceptible de porter une part de vérité qui transparaîtra dans des résultats expérimentaux non probants¹¹¹. La seconde faiblesse identifiée tient au design des expérimentations et à leur conduite fastidieuse. C'est ce qui amène les auteurs à proposer de nouveaux protocoles expérimentaux permettant de contrôler les variables susceptibles de perturber la validité externe ou interne de l'expérimentation, concepts que D. Campbell a contribué à introduire dans le champ scientifique¹¹².

Pendant les années 1950, la pratique de la méthode expérimentale par assignation aléatoire continue à se diffuser dans les milieux scientifiques et notamment dans les disciplines de l'éducation et de la psychologie¹¹³. Elle se développe aussi de manière fulgurante en médecine. Harry Marks dans son ouvrage « La médecine des preuves » soutient que ce succès est fortement lié à l'action d'une communauté de réformateurs thérapeutiques dont l'action fut d'ordre politique¹¹⁴. Il résume : « *que les essais cliniques soient le meilleur (ou le seul) instrument pour atteindre cet objectif (la pratique de la médecine s'appuyant sur des données probantes) est une question politique* »¹¹⁵. Cette communauté est unie par la croyance qu'une meilleure connaissance des effets et des usages des médicaments entraîne directement une meilleure pratique médicale. Ce développement intervient dans un contexte de scepticisme vis-à-vis des jugements personnels des praticiens sur les traitements à leur disposition. La méthode aléatoire, permet d'introduire un outil impersonnel d'intégrité scientifique¹¹⁶. La première utilisation de cette méthode en médecine est conduite par A. B. Hill en 1948. Son emploi en médecine incite à développer des protocoles contrôlant davantage la subjectivité de l'observateur - se manifestant ici par une préférence thérapeutique - en introduisant la pratique du double aveugle, à savoir que ni le médecin ni les patients ne savent qui bénéficie du traitement ou non. Dans ce domaine d'application, cette méthode

¹⁰⁹ Donald T CAMPBELL et Julian C STANLEY, *op. cit.* p. 3.

¹¹⁰ Traduction de l'auteur à partir de l'anglais : "We must expand our students' vow of poverty to include not only the willingness to accept poverty of finances, but also a poverty of experimental results.", *Ibid.* p. 3.

¹¹¹ "When one finds, for example, that competent observers advocate strongly divergent points of view, it seems likely on a priori grounds that both have observed something valid about the natural situation, and that both represent a part of the truth. The stronger the controversy, the more likely this is. Thus we might expect in such cases an experimental outcome with mixed results, or with the balance of truth varying subtly from experiment to experiment." *Ibid.* p. 3

¹¹² William R. SHADISH et Jason K. LUELLEN, « Donald Campbell: The Accidental Evaluator », dans Marvin C. ALKIN (dir.), *Evaluation roots: a wider perspective of theorists' views and influences*, 2nd ed, Los Angeles, SAGE Publications, 2004, p. 80-87, p.81.

¹¹³ Larry L. ORR, *Social experiments: evaluating public programs with experimental methods*, Thousand Oaks, Calif, Sage Publications, 1999, p. 24.

¹¹⁴ Harry MARKS, *La médecine des preuves : histoire et anthropologie des essais cliniques : 1900-1990*, Le Plessis-Robinson, Institut Synthélabo pour le progrès de la connaissance, 1999, p. 17.

¹¹⁵ *Ibid.* p.32

¹¹⁶ Harry MARKS, *op. cit.* p.18

est choisie pour sa force et sa simplicité de démonstration auprès d'un public non familiarisé avec les statistiques (ici les médecins). A.B. Hill explique : « *Ma compétence, si j'en ai une, était me semble-t-il d'offrir au clinicien quelque chose de simple qu'il pouvait comprendre (...) Si l'on avait commencé par des propos abscons, la réponse aurait été « allez vous faire voir » - et on en serait encore là* »¹¹⁷. La pratique de la méthode expérimentale par assignation aléatoire s'est peu à peu développée dans le cadre de *l'evidence-based medicine*, la médecine des preuves, nous reviendrons sur cette pratique plus longuement dans la deuxième partie. Ainsi, à la fin des années 1950, la méthode expérimentale par assignation aléatoire est finement théorisée et connaît déjà plusieurs domaines d'application dans l'aide à la décision. Néanmoins, cette pratique reste encore marginale et débattue dans les milieux scientifiques. C'est dans les années 1960 que cette méthode par son association avec l'expérimentation politique est définie comme le protocole des « vraies » expérimentations.

1.1.2. Le développement de l'utilisation de la méthode expérimentale aléatoire pour l'aide à la prise de décision, l'émergence de « la méthode la plus rigoureuse »

À partir des années 1960, la méthode expérimentale par assignation aléatoire devient donc un outil d'évaluation privilégié dans le cadre d'une démarche expérimentale en action publique. Elle est institutionnalisée comme « étalon or » des méthodes d'évaluation puis est appliquée aux grands projets réformateurs conduits par l'État fédéral américain. Suite à ces expérimentations, et au vu des difficultés de sa mise en œuvre mais aussi de l'ambiguïté des résultats produits, cette méthode suscite nombre de critiques d'acteurs de l'action publique et de scientifiques. Ces critiques amènent à redéfinir le périmètre de pertinence d'emploi de cette méthode pour l'évaluation de l'action publique. À l'issue de cette période, émerge une nouvelle pratique de l'expérimentation aléatoire dans l'action publique, guidée principalement par un souci de qualité des résultats produits afin de contribuer à l'apport de connaissances dans la science économique expérimentale.

1.1.2.1. La « société expérimentale »

À la fin des années 1950, puis principalement dans les années 1960, Donald Campbell théorise l'emploi de la méthode expérimentale par assignation aléatoire dans l'action publique. Il plaide pour une approche expérimentale de la réforme sociale la définissant comme : « *Une approche par laquelle nous essayons de nouveaux programmes conçus pour soigner des problèmes sociaux spécifiques, et par lesquels nous apprenons si ces programmes sont effectifs ou non, et que nous retenons, imitons, modifions ou abandonnons sur la base de leur effectivité apparente, à partir des multiples critères imparfaits dont nous disposons* »¹¹⁸. On observe d'ailleurs ici, par l'emploi du verbe « soigner », une référence explicite au modèle de *l'evidence-based medicine* et au développement, à cette époque, de l'utilisation de la méthode expérimentale dans ce domaine. D. Campbell enjoint les administrateurs publics, ainsi que les chercheurs en sciences sociales, à tendre si possible vers la conduite de « véritables expérimentations » par la conduite de la « meilleure méthode », c'est-à-dire avec des groupes de contrôle au

¹¹⁷ Cité dans *Ibid.* p.211

¹¹⁸ Traduction de l'auteure à partir de l'anglais: "an approach in which we try out new programs designed to cure specific social problems, in which we learn whether or not there programs are effective, and in which we retain, imitate, modify, or discard them on the basis of apparent effectiveness on the multiple imperfect criteria available" Donald T. CAMPBELL, « Reforms as Experiments », (1969) 24-4 *American Psychologist* 409- 429, 71.

sein desquels les individus sont répartis de manière aléatoire¹¹⁹. Conscient des obstacles de cette pratique dans ce domaine d'application, il invite au développement de « *postures politiques et idéologiques permettant de rendre possible le tirage aléatoire à ce niveau* »¹²⁰. Il propose plusieurs utilisations de l'expérimentation dans la conduite de l'action publique afin qu'elle soit acceptée. D'abord dans le cadre de programmes pilotes, dont on doit juger du succès avant de les généraliser, ensuite, dans le cadre de la diffusion de nouvelles réformes, petit à petit, de territoire en territoire, ou par le biais de loteries dans le cas de ressources limitées. Il souligne néanmoins que l'expérimentation en sciences sociales, de par la difficulté d'isoler l'expérimentation des facteurs extérieurs, nécessite d'autant plus de réplication d'expérimentations pour en consolider les résultats. C'est donc dans le contexte de la rencontre entre la recherche en sciences sociales et l'administration sociale que l'expérimentation aléatoire a été définie comme la « véritable » manière d'expérimenter¹²¹.

Au début des années 1960, la méthode expérimentale par assignation aléatoire se propage peu à peu comme une méthode d'évaluation couplée avec la pratique d'expérimentations sociales en action publique. Elle se diffuse d'abord de manière marginale, à partir de 1961, dans le champ de la justice (*Manhattan bail bond project*), mais c'est dans la deuxième moitié des années 1960, qu'elle connaît un essor considérable pour l'évaluation de projets de grande ampleur. Dans un contexte de remise en cause des nouvelles interventions de l'État providence, sous la présidence de L. B. Johnson, la méthode par assignation aléatoire a été mobilisée pour apporter une démonstration probante de l'efficacité de ces dispositifs de lutte contre la pauvreté. Un des projets phares de ces réformes est le « *New jersey income maintenance* », une intervention soutenue par les économistes de l'époque, qui consiste à mettre en place un revenu minimum par le biais d'un impôt négatif¹²². L'enjeu pour les agences fédérales est de convaincre les membres du Congrès que cette mesure n'allait pas inciter les bénéficiaires à arrêter la recherche d'emploi. Il est intéressant de constater que cette entrée de la conduite d'expérimentations par le domaine de l'emploi sera aussi le secteur d'entrée des premières expérimentations en France. Une piste d'explication de cette tendance est l'intérêt pour cette méthode des économistes statisticiens dans le domaine de l'emploi. C'est en effet, sous la proposition d'Heather Ross, une doctorante du MIT, l'*Office of Economic Opportunity* (l'agence de lutte contre la pauvreté de L. Johnson) choisit d'évaluer ce projet par la méthode aléatoire¹²³. Cette expérimentation à grande échelle est lancée en 1968 et porte sur environ 1 300 foyers¹²⁴. Cette expérimentation s'est déroulée sur une durée de 8 ans pour un coût de 132,7 millions de dollars¹²⁵. Les premiers résultats ont été connus à partir de 1971, les derniers en 1978. Les premiers résultats sont interprétés comme positifs sur l'impact de ce revenu sur la recherche de travail¹²⁶. Cependant, cette démonstration est remise en cause par les membres du

¹¹⁹ *Ibid.* p.74.

¹²⁰ Traduction de l'auteur à partir de l'anglais : "We need to develop the political postures and ideologies that make randomization at these levels possible" Donald T. CAMPBELL, *op. cit.* p.94.

¹²¹ Trudy DEHUE, « Establishing the experimenting society: The historical origin of social experimentation according to the randomized controlled design », *op. cit.* p. 286.

¹²² Cette intervention est notamment inspirée par les travaux de M. Friedman En 1962 *Capitalism and Freedom* comme le souligne Guillaume ALLEGRE, *L'expérimentation sociale des incitations financières à l'emploi, questions méthodologiques et leçons des expériences nord-américaines*, Document de travail de l'OFCE n°2008-22, 2008, p. 11.

¹²³ Larry L. ORR, *op. cit.* p. 1.

¹²⁴ Heather L. ROSS, *Testing Experimentation: Income Maintenance and Social Policy*, Cambridge, Mass, Edited by James G. Abert and Murray Kamrass.Ballinger, Social experiments and social program evaluation; proceedings of the Washington Operations Research Council Symposium., 1974.

¹²⁵ Eric MONNIER, *Evaluations de l'action des pouvoirs publics*, Paris, Economica, 1992, p. 40.

¹²⁶ Steven LEVITT et John LIST, *Field Experiments in Economics: the Past, the Present, and the Future*, National Bureau of Economic Research, Cambridge MA, 2008, p. 13.

Congrès, qui s'appuient sur les difficultés de la mise en œuvre du protocole, et notamment l'abandon de l'expérimentation par 22% des participants, pour en discréditer les résultats¹²⁷. Du fait de la faible probité des résultats, cette expérimentation est déclinée sur d'autres territoires, dans l'Iowa et la Caroline du Nord, à Seattle et Denver ou à Gary sur différents publics cibles, avec divers seuils d'accès à la prestation, et des traitements plus importants, afin d'augmenter la visibilité des effets et la validité externe des résultats¹²⁸. Au final, seuls les résultats de l'expérimentation de Seattle et Denver sont jugés significatifs et démontrent un retour à l'emploi moins important sur le groupe test, mais ne permettent pas de monter en généralisation du fait de la particularité de la prestation sur ce territoire et le manque de données significatives issues des autres expérimentations¹²⁹. Ces résultats sont source d'interprétations variées puisque les scientifiques les ayant conduites estiment avoir démontré qu'il y avait peu d'effets négatifs à ce type d'intervention, alors que la presse a souligné leur effet désincitatifs vis-à-vis de la recherche d'emploi¹³⁰. Cette première expérimentation de grande ampleur a été un échec politique. Mise en place principalement pour mobiliser un soutien politique plus que pour en mesurer l'effet, l'ambiguïté de ces résultats fournit surtout des arguments à ses opposants conservateurs¹³¹. Plusieurs grands projets ont été évalués sur le même modèle jusqu'au milieu des années 1980¹³².

La pratique expérimentale se diffuse dans d'autres agences fédérales comme le Ministère du Logement et de l'Aménagement urbain (*Housing and urban development*) ainsi que le celui du travail. La méthode expérimentale est promue comme la méthode expérimentale à privilégier pour l'évaluation des programmes publics par le ministère de la Santé qui déclare : « *dans la mesure du possible, utilisez l'assignation aléatoire comme méthode d'évaluation* »¹³³. On observe alors un changement de la logique expérimentale au cours de ces années, E. Monnier évoque le passage d'une période de grandes ambitions à celle des petits pas, de 1975 à 1983, et d'un réformisme prudent depuis 1983. Il explique notamment que les nouvelles expérimentations sont plus courtes, d'une durée de 1 à 3 ans et leur coût réduits de moitié¹³⁴. En effet, à partir des années 1980 les expérimentations deviennent plus ciblées, moins ambitieuses, mais plus nombreuses¹³⁵. Cette période correspond à l'accession de R. Reagan au pouvoir et des coupes budgétaires importantes dans les fonds alloués à l'évaluation, qui de 1980 à 1984 sont amputés de 37%, et dont le personnel affecté à l'évaluation en interne est aussi diminué de moitié de 1980 à 1988¹³⁶. À partir des années 1970 et dans le courant des années 1980, les expérimentations sociales sont aussi conduites par des fondations privées non gouvernementales dont le but affiché est de conseiller l'action publique de façon non partisane. La Rand Corporation, créée par la fondation Ford en 1948 finance des expérimentations sur plus de 65 000 bénéficiaires, et le MDRC (*Manpower*

¹²⁷ Eric MONNIER, *op. cit.*

¹²⁸ G. Allègre dans son article propose un tableau récapitulatif des différents protocoles mis en place dans ces déclinaisons Guillaume ALLEGRE, *op. cit.* p. 13.

¹²⁹ Gary BURTLESS, « The Case for Randomized Field Trials in Economic and Policy Research », (1995) 9-2 *The Journal of Economic Perspectives* 63- 84, 76.

¹³⁰ Guillaume ALLEGRE, *op. cit.* p. 17.

¹³¹ Larry L. ORR, *op. cit.* p. 25.

¹³² Pour consulter des tableaux récapitulatifs sur les expérimentations de cette période voir en anglais l'article de Burtless et en Français l'ouvrage d'E. Monnier Gary BURTLESS, « The Case for Randomized Field Trials in Economic and Policy Research », (1995) 9-2 *The Journal of Economic Perspectives* 63- 84; Eric MONNIER, *op. cit.*

¹³³ Traduction de l'auteur à partir de l'anglais « *to the maximum extent feasible, use random assignment as an evaluation methodology* » Larry L. ORR, *op. cit.* p. 26.

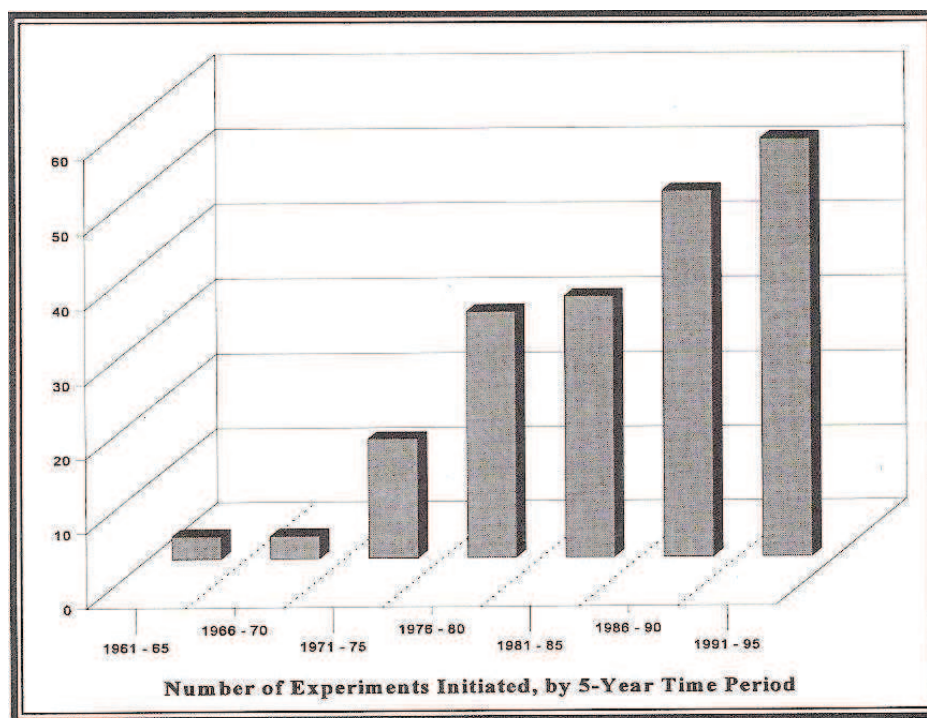
¹³⁴ Eric MONNIER, *op. cit.*

¹³⁵ Gary BURTLESS, *op. cit.* p. 63.

¹³⁶ William R SHADISH, Thomas D COOK et Laura C LEVITON, « Social Program Evaluation: Its History, Tasks, and Theory », dans *Foundations of program evaluation: theories of practice*, Newbury Park, Calif., Sage Publications, 1991, p. 19- 35 à la page 27.

Demonstration Research Corporation), créé par la même fondation en 1974, fait de la conduite d'expérimentations sa tâche principale¹³⁷.

Graphique 1 : Nombre d'expérimentations sociales évaluées par assignation aléatoires initiées entre 1961 et 1995 (source : Greenderg et Shroder 1997¹³⁸)



Lecture : le graphique montre le nombre d'expérimentations aléatoire dans le domaine de l'intervention sociale initiée par tranche de 5 ans. Le nombre total d'expérimentations initiées sur toute la période est calculé en additionnant chacune des tranches. Au total, selon ce recensement en 1980, 55 expérimentations avaient été initiées et 195 au total en 1995.

Alors que les grandes expérimentations proposent une diversité d'interventions sur un large territoire, permettant d'établir un effet moyen selon une variété de traitement, ces nouvelles expérimentations sont caractérisées comme des « boîtes noires », au sens où chaque traitement testé représente une intervention unique dont il est difficile ensuite d'extrapoler les résultats¹³⁹. Ces expérimentations de deuxième génération testent non plus des réformes radicales, mais des changements marginaux sur des programmes existants. Sur la période de 1962 à 1974, 80% des expérimentations aléatoires testent de nouveaux programmes et seulement 33% à partir de 1983¹⁴⁰. Jusqu'en 1994, l'expérimentation par tirage aléatoire a mobilisé plus d'un milliard de dollars¹⁴¹.

¹³⁷ « *The RAND Corporation is a nonprofit institution that helps improve policy and decision making through research and analysis* », Mission définie sur leur page internet <http://www.rand.org/about/history.html>, et le slogan du MDRC est « *building knowledge to improve social policy* », Trudy DEHUE, « Establishing the experimenting society: The historical origin of social experimentation according to the randomized controlled design », *op. cit.* p. 295.

¹³⁸ Larry L. ORR, *op. cit.* p. 26.

¹³⁹ Gary BURTLESS, *op. cit.* p. 67.

¹⁴⁰ Steven LEVITT et John LIST, *Field Experiments in Economics: the Past, the Present, and the Future*, National Bureau of Economic Research, Cambridge MA, 2008, p. 15.

¹⁴¹ Gary BURTLESS, *op. cit.* p. 63.

Au cours de ces années, les scientifiques se désinvestissent de cette entreprise suite aux différents débats et aux critiques à l'encontre de la scientificité des résultats produits. Ces débats ont été la source des différents aménagements de la pratique expérimentale et marquent aussi la naissance de la formalisation d'une nouvelle réflexion sur l'adaptabilité des méthodes scientifiques à l'environnement de l'action publique. À l'issue de cette période, au début des années 1990, l'expérimentation aléatoire est jugée comme une méthode non adaptée à l'évaluation des politiques publiques¹⁴².

Nous proposons de revenir maintenant sur les critiques de cette méthode puisqu'elles participeront à l'émergence d'une nouvelle pratique expérimentale à la fin des années 1990 et qu'elles contribueront à structurer le champ de l'évaluation des politiques publiques aux États-Unis et à repenser l'association des chercheurs et de l'action publique.

1.1.2.2. Le débat scientifique post utilisation : repenser l'évaluation au-delà de l'application de méthodes scientifiques à l'action publique

Ces premières utilisations de la méthode expérimentale pour éclairer la décision publique ont été la cible de critiques émanant des acteurs de l'action publique ainsi que des milieux scientifiques. Ces critiques s'organisaient en deux arguments, le manque d'adaptabilité de cette méthode à l'action publique et les failles de sa démonstration scientifique.

Les milieux scientifiques furent très mobilisés par cette activité d'évaluation via une rencontre de l'offre scientifique et de la demande gouvernementale. D'une part l'État fédéral ne disposait pas en interne d'un nombre d'experts suffisant ou assez spécialisés dans l'étude des phénomènes sociaux, d'autre part les doctorants en sciences sociales étaient en surnombre dans les années 1970 par rapport aux postes universitaires et l'évaluation devenait une carrière alternative faisant appel à leurs compétences d'observation et d'enquête¹⁴³. Les décennies 1960 et 1970 marquent les prémices de l'application à grande échelle des méthodes scientifiques des sciences sociales à l'action publique par la conduite d'expérimentations. Plusieurs chercheurs reconnaissent leur perception « naïve » quant à la conduite de l'action publique et de la décision politique ainsi que sur la possibilité de susciter un apprentissage politique par la seule application des méthodes scientifiques¹⁴⁴. Ces expériences ont constitué la source d'une abondance de réflexions dans les années 1980 et 1990 sur l'adaptabilité des méthodes scientifiques à l'évaluation de programmes publics.

Cette première période de théorisation et d'expériences d'application de la méthode expérimentale aléatoire en évaluation des politiques publiques a rencontré d'abord un engouement politique, suivi par de nombreuses critiques suite à la nature des résultats et à leur faible utilité pour éclairer l'action publique¹⁴⁵. Un décalage important est repéré entre, les postulats sur lesquels repose le protocole expérimental et le terrain de l'expérimentation. Les travaux d'A. Wildavsky et J. Pressman, notamment sur les

¹⁴² Trudy DEHUE, « Establishing the experimenting society: The historical origin of social experimentation according to the randomized controlled design », *op. cit.* p. 195.

¹⁴³ William R. SHADISH, Thomas D. COOK et Laura C. LEVITON, « Social Program Evaluation: Its History, Tasks, and Theory », dans *Foundations of program evaluation: theories of practice*, Newbury Park, Calif., Sage Publications, 1991, p. 19-35 p. 25.

¹⁴⁴ C. Weiss explique cette entreprise : « *this book is dedicated to overcoming the king of naivete that I began with* » Carol H. WEISS, *Evaluation research: methods for assessing program effectiveness*, coll. Prentice-Hall methods of social science series, Englewood Cliffs, N.J., Prentice-Hall, 1972, p. 13.

¹⁴⁵ Eric MONNIER, *op. cit.*

expérimentations conduites à partir de 1965 par l'agence du développement économique (*Economic Development Agency*) financée par les fonds fédéraux, soulignent la plasticité d'un projet au cours de sa mise en œuvre, sujette aux décisions locales et donc à des modifications d'objectifs, de publics, de moyens par rapport à sa programmation originale¹⁴⁶. D'ailleurs, une des motivations du développement de l'évaluation sur ces projets a été aussi de contrôler ces derniers afin qu'ils correspondent à leur programmation¹⁴⁷. La méthode aléatoire apparaît donc comme peu propice à l'évaluation d'expérimentation au sein de projets innovants voués à être modifiés et à faire l'objet d'ajustements locaux¹⁴⁸. C. Weiss, une des plus prolifiques théoriciennes de la recherche évaluative, a essayé d'employer la méthode aléatoire à ses débuts, et témoigne : « *Ce qui semble élémentaire en théorie s'avère en pratique constituer une entreprise exigeante. Les programmes sont loin d'être aussi ordonnés [neat] et accommodants que l'attend l'évaluateur. Pas plus que les circonstances extérieures ne sont aussi passives et sans conséquence qu'il le souhaiterait* »¹⁴⁹.

Tout un courant de chercheurs en évaluation confirme l'importance de prendre en compte l'environnement politique de l'évaluation¹⁵⁰. Ils soulignent que la meilleure méthode d'évaluation n'est pas la plus scientifique, mais la plus à même de s'adapter à l'environnement politique, de le prendre en compte et de lui procurer des informations qu'il est susceptible d'utiliser¹⁵¹. Les préoccupations sur la scientificité des méthodes, sont relayées par des préoccupations sur le contexte de la pratique d'évaluation et la place des résultats d'évaluation dans un contexte politique¹⁵². Ces réflexions donnent naissance à de nouvelles démarches évaluatives destinées principalement à favoriser l'apprentissage des acteurs de l'action publique comme l'évaluation basée sur la théorie (*theory-based evaluation*) et celle basée sur son utilisation (*utilization-focused evaluation*)¹⁵³. Même le théoricien de la « société expérimentale » estime, dès la fin des années 1970, que cette méthode de démonstration de la causalité, bien qu'elle reste la plus rigoureuse à ses yeux, n'est pas adaptée à l'évaluation de programmes publics qui suscitent trop de biais non contrôlables dans leur conduite et particulièrement dans la conduite de programmes innovants. L'encadré ci-dessous reprend ses principaux arguments.

Encadré 2 : Critique de D. Campbell sur l'utilisation des expérimentations aléatoires pour l'évaluation des politiques publiques

Critique de D. Campbell de l'application de la méthode expérimentale aux interventions publiques

“The critics taking what I am calling the humanistic position are often well-trained in quantitative-experimental methods. Their specific criticisms are often well-grounded in the experimentalist's own framework: experiments implementing a single treatment in a single setting are profoundly ambiguous as to what caused what; there is a precarious rigidity in the measurement system, limiting recorded

¹⁴⁶ Jeffrey L PRESSMAN et Aaron B WILDAVSKY, *Implementation.*, Berkeley, Calif., University of California, 1984.

¹⁴⁷ William R SHADISH, Thomas D COOK et Laura C LEVITON, *op. cit.* p. 23.

¹⁴⁸ Peter H. ROSSI et Howard E. FREEMAN, *op. cit.* p. 290.

¹⁴⁹ Traduction de l'auteur à partir de l'anglais : « *what looks elementary in theory turns out in practice to be demanding entreprise. Programs are nowhere near as neat and accommodating as the evaluator expects. Nor are outside circumstances as passive and unimportant as he might like* » Carol H. WEISS, *Evaluation research: methods for assessing program effectiveness*, *op. cit.* p. 25.

¹⁵⁰ Dennis James PALUMBO (dir.), *The Politics of program evaluation*, coll. Sage yearbooks in politics and public policy, v. 15, Newbury Park, Calif, Sage Publications, 1987.

¹⁵¹ Carol H WEISS, « Where Politics and Evaluation Research Meet », (1993) 14-1 *Evaluation Practice* 93-106.

¹⁵² Lee J. CRONBACH, *Toward reform of program evaluation*, San Francisco, Jossey-Bass Publishers, 1980.

¹⁵³ Pour un retour historique, sur ce courant et ses principes voir notre article ou l'ouvrage de M. Q. Patton Agathe DEVAUX-SPATARAKIS, « L'évaluation “basée sur la théorie” entre rigueur scientifique et contexte politique », (2014) 31-1 *Politiques et Management Public* 75-91; Michael Quinn PATTON, *Utilization-focused evaluation: the new century text*, Thousand Oaks; London; New Delhi, Sage Publications, 1997.

outcomes to those dimensions anticipated in advance; process is often neglected in an experimental program focused on the overall effect of a complex treatment, and thus knowing such effects has only equivocal implications for program replication or improvement; broad-gauge programs are often hopelessly ambiguous as to goals and relevant indicators; change of treatment program during the course of an ameliorative experiment, while practically essential, make input-output experimental comparisons uninterpretable; social programs are often implemented in ways that are poor from an experimental design point of view; even under well-controlled situations, experimentation is a profoundly tedious and equivocal process; experimentation is too slow to be politically useful; etc. All these are true enough, often enough to motivate a vigorous search for alternatives¹⁵⁴.

Most of what most governments offer in the name of "reforms" and "new programs" are symbolic gestures designed to indicate government awareness of problems and sympathetic intention, rather than serious efforts to achieve social change...Due to competing needs and limited budgets, most of the few genuine novelties get drastically underfunded and have no chance of producing demonstrable effects. The legislative requirement that such so-called programs be scientifically evaluated becomes just more empty rhetoric, a token to indicate that the money is being spent responsibly¹⁵⁵.

L'objet même de l'évaluation proposé par l'expérimentation aléatoire, se réduisant à l'identification de la causalité par l'estimation de l'impact est jugé insuffisant pour rendre compte de projets enchevêtrés dans des enjeux économiques, politiques et sociaux. Dans leur article sur D. Campbell, W. Shadish et J. Leullen concluent que *« finalement, le champ de l'évaluation a rejeté la « société expérimentale » de D. Campbell comme trop étroite et utopique, préférant une vision plus large du rôle de l'évaluation »*¹⁵⁶.

Le débat sur la pertinence de l'utilisation de la méthode aléatoire avait à l'époque mobilisé l'ensemble de la communauté émergente des évaluateurs, en témoigne l'organisation en 1981 d'une conférence conjointe de *The Evaluation Network* et *The Evaluation Research Society* (qui ont plus tard formé *l'American Evaluation Association*), intitulée *« Should the federal government mandate the use of experimental methods in evaluation ? »*¹⁵⁷. À l'issue de cette conférence, le périmètre d'application de la méthode aléatoire est redéfini selon la nature du projet évalué¹⁵⁸. Cette méthode d'évaluation doit être appliquée uniquement sur des projets dont les relations causales sont simples et peu nombreuses, avec un faible nombre de variables à mesurer, le programme doit être stable, et les autres variables exogènes faciles à contrôler. Elle doit aussi n'avoir aucune autre ambition que d'estimer l'impact de l'intervention. Dans leur ouvrage *Evaluation : a systematic approach*, paru au début des années 1990, P. Rossi et H. Freeman complètent ces conditions d'emploi¹⁵⁹. Ils estiment que cette méthode ne doit pas être utilisée au début du développement d'un programme, puisqu'il risque d'évoluer, qu'elle ne doit pas être utilisée si l'action publique a besoin d'informations rapidement, qu'elle porte sur des interventions dont les services ne peuvent pas être standardisés et que les risques de non-recours au programme ou de déperdition dans le suivi des cohortes ne peuvent pas être contrôlés en amont.

¹⁵⁴ Donald T CAMPBELL, « Assessing the impact of planned social change », (1979) 2 *Evaluation and Program Planning* 67- 90, 69.

¹⁵⁵ Donald T CAMPBELL, « Experiments as arguments », *op. cit.* p. 118.

¹⁵⁶ Traduction de l'auteure à partir de l'anglais *« Eventually, the field of evaluation rejected Campbell's Experimenting Society as too narrow and Utopian, pre-ferring a broader vision of the role of evaluation. »* William R. SHADISH et Jason K. LUELLEN, « Donald Campbell: The Accidental Evaluator », dans Marvin C. ALKIN (dir.), *Evaluation roots: a wider perspective of theorists' views and influences*, 2nd ed, Los Angeles, SAGE Publications, 2004, p. 80-87 p. 82.

¹⁵⁷ Nick L. SMITH et Paul R. BRANDON (dir.), *Fundamental issues in evaluation*, New York, Guilford Press, 2008, p. 5.

¹⁵⁸ Nick L. SMITH, « The feasibility and desirability of experimental methods in evaluation », (1981) 3 *Evaluation and Program Planning*.

¹⁵⁹ Peter H. ROSSI et Howard E. FREEMAN, *op. cit.* p. 290.

Cependant, malgré cette redéfinition de son périmètre d'utilisation, cette méthode reste toujours majoritairement considérée comme la meilleure méthode scientifique pour établir la causalité d'une intervention lorsqu'elle est applicable. Si son utilisation est moins répandue dans l'évaluation de l'action publique, elle reste néanmoins un idéal à atteindre en termes de rigueur de démonstration. L'article de P. Rossi et S. Wright témoigne de cette position : *"There is almost universal agreement among evaluation researchers that the randomized controlled experiment is the ideal model for evaluating the effectiveness of a public policy. If there is a Bible for evaluation, the Scriptures have been written by Campbell and Stanley (1966), along with a revised version by Cook and Campbell (1975). The "gospel" of these popular texts is that all research designs can be compared more or less unfavorably to randomized controlled experiments, departures from which are subject to varying combinations of threats to internal and external validity."*¹⁶⁰.

1.1.2.3. L'appropriation de la méthode expérimentale par assignation aléatoire par les économistes statisticiens appliqués

Pendant cette période de débat succédant aux premières utilisations de grande ampleur de la méthode expérimentale par assignation aléatoire pour évaluer des interventions publiques, les économistes statisticiens appliqués se saisissent de cette méthode. Nous comprenons par économistes statisticiens appliqués, les économistes ayant recours à des méthodes quantitatives et particulièrement aux méthodes statistiques et s'intéressant à l'estimation de l'impact des interventions publiques. Leur intérêt pour la méthode expérimentale par assignation aléatoire s'inscrit dans une période marquée par le développement de la micro-économie comportementaliste, alliant psychologie et économie dans l'étude des comportements humains - parfois considérées comme déviant du modèle de rationalité de *l'homo economicus* - face à des incitations variées. Le nouvel intérêt de ces chercheurs pour l'expérimentation dans la production de connaissances scientifiques contribue à faire émerger une troisième vague d'expérimentations aléatoires portant sur des interventions à plus petite échelle et susceptibles de tester ces mêmes théories de micro-économie.

Ces premières expériences suscitent aussi un débat important dans la communauté des économistes statisticiens appliqués américains. S'opposent d'une part, les partisans de la méthode expérimentale par assignation aléatoire et d'autre part, les partisans d'une démonstration non expérimentale ou économétrique¹⁶¹. Dans un article de 1986, R. Lalonde démontre les biais de sélection présents dans les autres méthodes non expérimentales en les comparant avec les résultats d'une expérimentation aléatoire sur la même politique de retour à l'emploi¹⁶². Cette démonstration contribue alors à diffuser cette méthode dans l'évaluation des programmes de formation ou de retour à l'emploi. J. Heckman et J. Smith, en 1995, remettent en cause la linéarité de l'interprétation des résultats entre l'idéal de la conduite expérimentale et la pratique qui peut être soumise à plusieurs biais d'échantillonnage, ou d'attrition (manque de suivi des cohortes)¹⁶³. Ces

¹⁶⁰ Peter H. ROSSI et S.R. WRIGHT, « Evaluation research: An assessment of theory, practice, and politics », (1977) 1 *Evaluation Quarterly* 5-52, p.13.

¹⁶¹ James J. Heckman et Jeffrey A. Smith, « Assessing the Case for Social Experiments », *The journal of Economic perspectives* 9, n° 2 (1995): p.85.

¹⁶² Robert J. LALONDE, « Evaluating the Econometric Evaluations of Training Programs with Experimental Data », (1986) 76-4 *The American Economic Review* 604-620.

¹⁶³ James J. HECKMAN et Jeffrey A. SMITH, *op. cit.*

auteurs s'opposent à la démonstration de R. Lalonde, ils estiment que les évaluations non expérimentales peuvent atteindre la composition de groupes non sujets au biais de sélections si la qualité des données d'origines est améliorée. S'appuyant sur les expérimentations en cours, J. Heckman critique l'optimisme de production de connaissance démesuré de cette période, ce qui n'est pas sans rappeler l'analyse de D. Campbell sur la première vague d'emploi de cette méthode en éducation¹⁶⁴. Cet auteur remet d'ailleurs en cause l'application même de cette méthode théorisée dans l'agronomie aux sciences sociales en ironisant : « *Des lopins de terre ne sont pas sujets à l'anticipation de leur traitement par les engrais, et ne peuvent pas non plus refuser d'être traités* »¹⁶⁵.

Au-delà de ces critiques liées à la mise en œuvre de la méthode expérimentale sur le terrain, cette méthode connaît une utilisation croissante en économie en laboratoires. S'appuyant sur les travaux de D. Rubin, les économistes en micro-économie comportementaliste identifient cette méthode comme la mieux à même d'estimer l'effet causal d'un traitement en contrôlant notamment le maximum de biais pouvant intervenir lors de la composition des groupes de comparaison¹⁶⁶. On observe au cours des années 1990, un développement spectaculaire du recours à la méthode expérimentale par assignation aléatoire en économie et particulièrement en micro-économie comportementaliste (*behavioral*), aussi bien en laboratoire que sur le terrain¹⁶⁷. L'étude de D. Carde et al., sur l'évolution du nombre de publications ayant recours à ces méthodes dans les cinq plus prestigieuses revues académiques en économie, révèle une augmentation notoire de ces travaux à partir du milieu des années 1990¹⁶⁸. Alors que de 1975 à 1994 seulement 12 expérimentations aléatoires de terrain avaient été publiées dans ces revues, de 1994 à 2010, 72 papiers publiés ont eu recours à cette méthode¹⁶⁹. En 2010, les expérimentations en laboratoires en économie composaient 9,3% des articles dans ces mêmes revues et les expérimentations de terrain 2,5%.

¹⁶⁴ James J. Heckman, « Randomization and Social Policy Evaluation », in *Evaluating Welfare and Training Programs*, par C. Manski et I. Garfinkel, Harvard University Press (Cambridge, MA, 1992).

¹⁶⁵ Traduction de l'auteur à partir de l'anglais: "*plots of ground do not respond to anticipated treatments of fertilizer, nor can they excuse themselves from being treated.*" citation de J. Heckman dans Steven LEVITT et John LIST, *Field Experiments in Economics: the Past, the Present, and the Future*, Cambridge MA, 2008, p. 16.

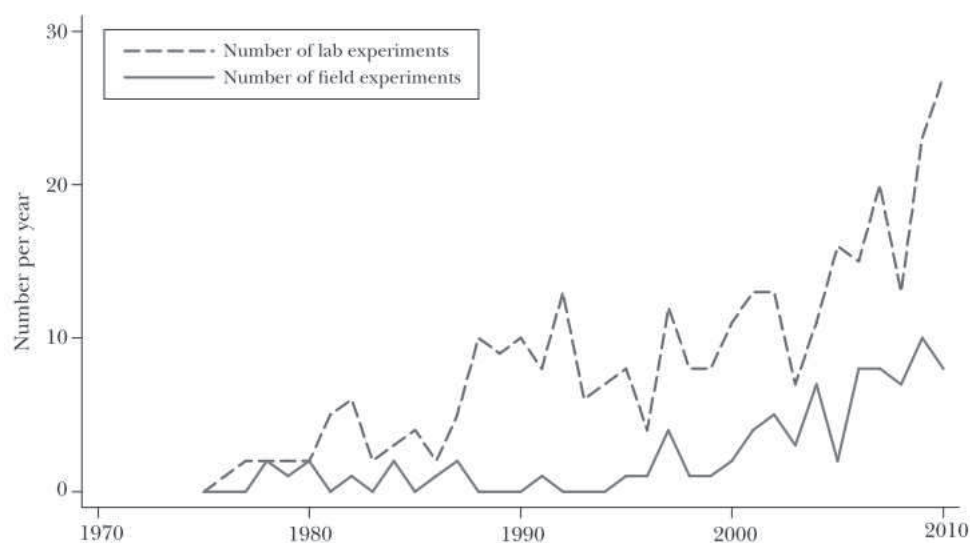
¹⁶⁶ Donald B. RUBIN, « Estimating causal effects of treatments in randomized and non-randomized studies », (1974) 66-5 *Journal of Educational Psychology* 688-701.

¹⁶⁷ Témoin de cet essor la réception du prix Nobel en 2002 par D. Kahneman tête de file de ce courant. Sa démarche à la rencontre de l'économie et de la psychologie est explicitée dans son ouvrage récemment traduit en français : Daniel KAHNEMAN et Raymond CLARINARD, *Système 1, système 2: les deux vitesses de la pensée*, Flammarion, 2012.

¹⁶⁸ Ces revues sont: *American Economic Review*, *Econometrica*, *Journal of Political Economy*, *Quarterly Journal of Economics* et *Review of Economic Studies*

¹⁶⁹ David CARD, Stefano DELLAVIGNA et Ulrike MALMENDIER, « The Role of Theory in Field Experiments », (2011) 25-3 *Journal of Economic Perspectives* 39-62, p.45.

Graphique 2 : Nombre d'expérimentation en laboratoire (laboratory experiments) et d'expérimentations de terrain (field experiments) publiées dans les cinq principales revues académiques en économie de 1975 à 2010



Note: The journals surveyed were the *American Economic Review*, *Econometrica*, the *Journal of Political Economy*, the *Quarterly Journal of Economics*, and the *Review of Economic Studies*.

Source : David CARD, Stefano DELLAVIGNA et Ulrike MALMENDIER, « The Role of Theory in Field Experiments », (2011) 25-3 *Journal of Economic Perspectives* 39-62, p.44.

Les expérimentations de terrain sont davantage orientées vers le questionnement de théories économiques et ont vocation à s'articuler avec la littérature produite par les expérimentations en laboratoire¹⁷⁰. On observe ainsi, à partir des années 2000, l'émergence d'un intérêt académique pour l'organisation d'expérimentations aléatoires de terrain en micro-économie. Cet engouement marque le retour des chercheurs dans cette pratique et le renouveau de la poursuite de leurs propres intérêts académiques dans la conduite d'expérimentation. Dans leur article sur l'évolution de la pratique expérimentale, Levitt et List notent cette évolution dans les années 2000, qu'ils caractérisent comme la troisième vague de l'expérimentation aléatoire de terrain. Les expérimentations sont de plus en plus imprégnées par les exigences et les intérêts de démonstration des économistes y participant. Elles sont réalisées sur des interventions de plus petite échelle, ne s'intéressent plus à l'évaluation d'interventions clés, de grands projets, mais plutôt à des dispositifs permettant d'observer des mécanismes débattus dans la micro-économie¹⁷¹. Elles sont aussi davantage conduites de manière opportuniste, sur des expérimentations « naturelles », au sens où le fait de priver un groupe de potentiels bénéficiaires n'est pas imposé par le protocole expérimental, mais par un mode de développement du projet ou une insuffisance de fonds. Il s'avère aussi que ces expérimentations se caractérisent par une ambition théorique plus importante et qu'elles sont plus orientées afin de tester des théories économiques que destinées particulièrement à parler à des décideurs publics¹⁷². Comme le souligne A. Deaton, le projet évalué devient alors une incarnation de la théorie économique à redéfinir et ne compose plus un objet d'évaluation en lui-même¹⁷³. De fait, la conduite de ce nouveau type d'expérimentations requiert de nouvelles compétences chez les chercheurs, celle de

¹⁷⁰ Steven LEVITT et John LIST, *Field Experiments in Economics: the Past, the Present, and the Future*, National Bureau of Economic Research, Cambridge MA, 2008, p. 2008.

¹⁷¹ Steven LEVITT et John LIST, *op. cit.* p. 19.

¹⁷² *Ibid*

¹⁷³ Angus DEATON, « Instruments, Randomization, and Learning about Development », (2010) 48 *Journal of Economic Literature* 424- 455, 451.

reconnaitre les opportunités d'expérimentations, de connaître assez bien le design de l'expérimentation, les connaissances théoriques en économie interrogées par ce cas et les compétences interpersonnelles pour organiser les relations complexes à l'œuvre entre les parties prenantes de l'expérimentation¹⁷⁴.

Ainsi, depuis la première expérimentation du New Jersey income tax jusqu'à l'année 2004, M. Schroder et D. Greenberg recensent la conduite de 235 expérimentations aléatoires appliquées à l'action publique aux États-Unis¹⁷⁵. La diffusion de l'utilisation de cette méthode au cours du XXe siècle et les débats dont elle fut l'objet nous renseignent sur les différentes théories véhiculées par cette dernière ainsi que son domaine d'application jusqu'au début des années 2000. L'expérimentation aléatoire appliquée à l'évaluation des politiques publiques est traversée par une tension entre le processus de production de preuve guidé par l'idéal scientifique du contrôle de la subjectivité de l'évaluateur la pratiquant et des variables du terrain d'une part, et les nécessaires aménagements de protocoles nécessités par son application à l'action publique d'autre part. Elle prend des formes différentes selon les acteurs qui la conduisent, et l'équilibre qu'elle organise entre objectifs scientifiques et politiques est variable sur la période étudiée.

Les années 1960, marquant son premier emploi de grande ampleur, témoignent d'une espérance des chercheurs d'appliquer cette méthode selon ses standards scientifiques à l'action publique afin de nourrir à la fois les orientations de l'action publique et de conforter des théories économiques. Les débats suite à ces premières utilisations soulignent les difficultés de garantir la rigueur scientifique sur le terrain et amènent les chercheurs à préférer des méthodes plus adaptées à l'action publique et à restreindre l'utilisation de la méthode aléatoire à un type d'intervention très restreint. Elle est désinvestie dans les années 1980 par les instances de gouvernement et les chercheurs et devient l'outil d'organisations non gouvernementales pour mettre à l'épreuve l'efficacité de dispositifs d'action publique moins ambitieux, sans chercher à tester des théories scientifiques particulières sur un format adapté aux décideurs publics. Son appropriation par les économistes statisticiens, à partir des années 1990, la replace en tant qu'outil de production scientifique amené à tester et affiner des théories économiques. Elle propose alors, une réconciliation de ces deux logiques en abandonnant l'ambition d'organiser l'action publique selon ses standards, mais de guetter des opportunités expérimentales, c'est-à-dire des dispositifs d'action publique qui dès leur conception constituent des terrains propices à l'emploi de cette méthode. Dans cette nouvelle logique, la méthode expérimentale n'est plus une méthode d'évaluation dont la vocation principale est d'estimer l'efficacité de certaines modalités d'interventions publiques en vue de leur généralisation, mais devient une méthode scientifique pour laquelle l'action publique est un terrain permettant de tester des théories qui à termes seront susceptibles d'être source de connaissances pour les acteurs publics.

Cette nouvelle conceptualisation de l'expérimentation aléatoire marque une nouvelle période de développement du recours à cette méthode, à partir du milieu des années 2000. Son recours est soutenu par les micro-économistes statisticiens souhaitant étendre leur terrain d'expérimentation à de nouveaux domaines d'intervention publique. Néanmoins, afin de garantir sa bonne mise en œuvre, elle requiert une modification du « terrain » politique afin qu'il soit plus favorable à la démarche expérimentale et qu'il

¹⁷⁴ Steven LEVITT et John LIST, *op. cit.* p. 20.

¹⁷⁵ Si l'on met en perspective ces chiffres avec le bilan de 1995 dressé par les mêmes auteurs on compte 40 expérimentations initiées en 9 ans soit plutôt une baisse des expérimentations sociales à cette période par rapport au début des années 1990. David H. GREENBERG et Mark SHRODER, *The digest of social experiments*, 3rd ed, Washington, D.C, Urban Institute Press, 2004.

n'entrave pas son protocole scientifique. C'est là, l'ambition de la transformation proposée par le mouvement de *l'Evidence-based policy* auquel nous proposons de nous intéresser à présent.

1.2. La méthode expérimentale par assignation aléatoire comme instrument de *l'Evidence-Based Policy*

Depuis les années 2000, plusieurs gouvernements nationaux ou organisations non gouvernementales internationales promeuvent le développement du recours à des travaux scientifiques pour éclairer la décision politique. Ce courant a été appelé *l'evidence-based policy* (EBP) et fait explicitement référence à *l'evidence-based medecine* (médecine des preuves) développée 20 ans auparavant¹⁷⁶. Si de prime abord il semble que ce courant appelle à une plus forte influence de la science sur les acteurs de l'action publique, un regard plus détaillé sur ses incarnations pratiques révèle aussi que ce courant organise une plus forte influence des décideurs publics sur la production scientifique. L'institutionnalisation du recours à la preuve scientifique pour orienter l'action publique nécessite au préalable de déterminer ce qui sera considéré comme données probantes. L'étude de la diffusion de ce courant permet de mettre à jour le rôle déterminant des praticiens de l'expérimentation aléatoire (chercheurs ou organisations non gouvernementales) dans la diffusion et formalisation de l'EBP. Selon les configurations nationales et leur accès aux instances gouvernementales, ils parviennent avec plus ou moins de succès à établir l'expérimentation aléatoire comme l'instrument privilégié de l'EBP.

L'étude des différentes incarnations de l'EBP, et notamment celle comparée de sa diffusion en Grande-Bretagne aux États-Unis et au sein des organisations internationales de développement, témoigne de l'influence des acteurs de sa promotion dans l'élaboration d'un cadrage cognitif amenant à favoriser certaines modalités de production de preuves plutôt que d'autres. Nous proposons de revenir ici sur les liens existants entre *l'Evidence-Based Policy* et la méthode expérimentale par assignation aléatoire.

Dans un premier temps, il est nécessaire de définir ce mouvement et d'identifier ce qui le distingue des précédentes entreprises visant à favoriser l'interaction des sciences sociales et de l'action publique. L'originalité de l'EBP est de proposer une double entreprise. D'une part, elle tente de formaliser un système de management de la connaissance issu de travaux scientifiques au sein des instances de l'action publique et, d'autre part, elle incite les chercheurs à produire davantage de travaux sur l'action publique correspondant à des standards scientifiques définis par les instances publiques. L'étude de la diffusion de ce mouvement montre comment ce modèle standardisé emprunté à d'autres secteurs d'intervention connaît des variations dans ses incarnations nationales selon la priorité donnée à la réforme du management de la connaissance au sein des instances d'action publique ou à la production de connaissances scientifiques. La place accordée à l'expérimentation aléatoire dans la mise en place de l'EBP ouvre une nouvelle période de débats sur cette méthode au sein de la communauté scientifique et des évaluateurs. L'étude de ceux-ci permet de mettre à jour les théories véhiculées par cette nouvelle pratique de la méthode expérimentale autant vis-à-vis de la démonstration de la causalité que sur la nature de l'intervention publique.

¹⁷⁶ Harry MARKS, *op. cit.*

1.2.1. Les incarnations de *l'Evidence-Based Policy*, les tensions entre modèle standardisé et contextes nationaux

La notion d'*evidence-based policy* est un objet de recherche difficilement appréhendable. C'est avant tout une rhétorique, un objectif de réforme de l'action publique orientée vers l'efficacité, mais dont les instruments sont variables selon les contextes culturels et institutionnels. Cette plasticité donne à voir les dynamiques d'interaction entre les acteurs de production de travaux scientifiques portant sur l'action publique et les acteurs de l'action publique qui transparaissent dans l'élaboration des instruments d'action publique devant contribuer à réformer les pratiques politiques. Néanmoins, cette notion fait aussi référence dans sa formulation au modèle de *l'evidence-based medicine* (EBM), qui a déjà développé ses propres instruments, à travers l'identification des données probantes, leur synthèse et leur prise en compte. Alors qu'au début de son développement en Grande-Bretagne, les promoteurs de l'EBP ont cherché à élaborer leur propre modèle de pratique adapté à l'action publique, il semble qu'à la fin des années 2000 l'EBP prenne davantage la forme d'un modèle standardisé, proposant un transfert technologique des instruments de l'EBM à l'action publique. La méthode expérimentale par assignation aléatoire est alors identifiée comme l'instrument clé de cette démarche.

1.2.1.1. *L'evidence-based policy*, l'institutionnalisation de la production et de la prise en compte de la preuve scientifique

Le terme d'*evidence-based policy* peut prêter à confusion, puisqu'il ne s'agit pas d'un courant défendant une décision publique s'appuyant exclusivement sur la preuve scientifique. Le but des défenseurs de ce courant est surtout de faciliter d'une part, l'accès des décideurs publics à des travaux scientifiques pouvant éclairer la programmation politique, et d'encourager d'autre part, la production de travaux scientifiques apportant des éclairages pour l'amélioration de l'action publique¹⁷⁷. Il serait donc plus pertinent de parler du développement de l'influence des travaux scientifiques en politique, *evidence-influenced policy*, voire juste de la connaissance de ces travaux par les décideurs publics, *evidence-aware policy*¹⁷⁸. Néanmoins, le terme *evidence-based* étant commun à une vingtaine de pratiques professionnelles s'appuyant sur la preuve, comme *l'evidence-based management* ou *evidence-based education* par exemple, cette appellation, bien que trompeuse, a été maintenue¹⁷⁹. En outre, cette communauté d'appellation l'inscrit dans une filiation avec ces pratiques dans d'autres domaines et suggère l'emploi d'instruments standardisés pouvant être importés d'un secteur à l'autre.

La traduction du terme *evidence-based policy* en français est problématique. Certains, s'appuient sur la traduction d'autres pratiques d'*evidence-based*, et comme *l'evidence based medicine* avait été traduit par « médecine des preuves », privilégient l'appellation

¹⁷⁷ Catherine LAURENT, « Pourquoi s'intéresser à la notion d'"evidence-based policy"? », *Revue Tiers Monde*, 2009, p.861.

¹⁷⁸ Sandra M NUTLEY, Peter C SMITH et H. T. O DAVIES, « Introducing evidence-based policy and practice in public services », dans *What works? : evidence-based policy and practice in public services*, The Policy Press, 2000, p. 11.

¹⁷⁹ Russell GERSTEN et John HITCHCOCK, « What is Credible Evidence in Education? The Role of What Works Clearinghouse in Informing the Process », dans Christina A CHRISTIE, Melvin M MARK et Stewart I DONALDSON (dir.), *What counts as credible evidence in applied research and evaluation practice?*, Los Angeles, SAGE, 2009, p. 78- 96 à la page 5.

« politique des preuves »¹⁸⁰. D'autres, se refusent à traduire cette notion pour ne pas la vider de son sens lors du procédé de traduction¹⁸¹. En effet, le terme *evidence* peut être traduit par le terme de « preuve » disponible, mais il désigne aussi un témoignage, un signe observable par les sens, il suggère une vérification empirique¹⁸². Le terme *policy* évoque en anglais la dimension rationnelle et technique de l'action publique, la politique publique orientée vers la résolution des problèmes publics. Ce qu'on traduit en français par les politiques publiques. Ce terme se distingue en anglais de la *politics* désignant les enjeux partisans, les luttes de pouvoir, et de la *polity* évoquant un système global d'organisation du gouvernement. Au vu de ce travail de définition, traduire cette expression par la politique basée sur la preuve en revient à perdre des dimensions importantes des termes anglais. Nous préférons, dans nos écrits, garder le terme anglo-saxon, d'autant plus qu'il n'existe aucune traduction officielle de cette expression employée par les acteurs politiques français.

Ce programme de réformes se détache des différentes initiatives précédentes de rapprochement entre la science et l'action publique à plusieurs égards. Comme toutes les précédentes initiatives de développement de l'influence de la science dans la conduite de l'action publique, le contexte de l'EBP est celui d'une perte de confiance des électeurs en l'action publique, ainsi qu'une volonté d'économie gouvernementale en temps de crise¹⁸³. Néanmoins, les solutions qu'elle propose sont différentes. Le processus de rationalisation de l'action publique pour en améliorer l'efficacité, n'est pas une tendance nouvelle, M. Foucault dans ses écrits sur la « gouvernementalité » repère son origine au XVIII^e siècle en France par l'administration de N. Delamare et Necker qui témoigne d'un exercice du pouvoir tourné vers une rationalisation et technicisation s'appuyant sur des appareils de gouvernements et un ensemble de savoirs¹⁸⁴. Cette volonté d'interférence de la science avec l'action publique est présente dès l'origine de la constitution de la discipline d'analyse des politiques publiques, puisqu'un de ses pionniers H. Lasswell appelait, dès les années 1950, à la formation des « sciences politiques de la démocratie » (*policy sciences of democracy*) qui doivent être « orientées vers la production de connaissance afin d'améliorer la pratique de la démocratie »¹⁸⁵. L'*evidence-based policy* s'inscrit dans ce mouvement, au sens où elle requiert d'éclairer l'action par des savoirs spécifiques et particulièrement scientifiques. Cependant, elle s'en distingue, au sens où elle ne s'inscrit pas dans le recours aux savoirs d'experts, mais à l'institutionnalisation d'une méthode systématique de la recherche de preuves appropriées pour éclairer l'action publique¹⁸⁶.

En effet, les pratiques dites « *evidence-based* » impliquent d'une part, l'identification par les organisations du secteur auquel elles s'appliquent de ce qui est considéré comme preuve, et d'autre part la mise en place de méthodologies et de pratiques transparentes

¹⁸⁰ Alain DESROSIERES, « La statistique, outil de gouvernement et outil de preuve », dans *L'argument statistique*, Paris, Presses de l'école des mines, 2008

¹⁸¹ Catherine LAURENT, *op. cit.* p. 860.

¹⁸² LE ROBERT et BUREAU MARCEL VAN DIJK, *Le grand Robert & Collins dictionnaire français-anglais, anglais-français*, Dictionnaires Le Robert-Sejer, 2008.

¹⁸³ William SOLESBURY, *Evidence Based Policy: Whence it Came and Where it's Going*, ESRC UK Centre for Evidence Based Policy and Practice, Queen Mary, University of London, octobre 2001, p. 6

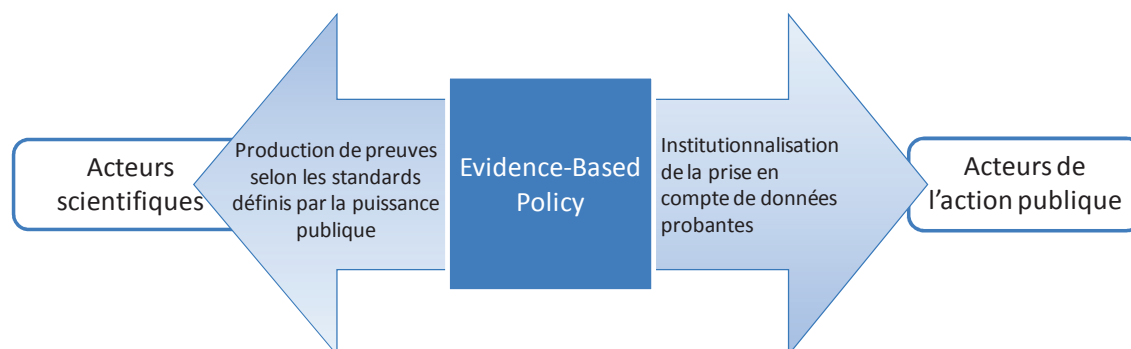
¹⁸⁴ Patrick LE GALES et Pierre LASCOUMES, « L'action publique saisie par ses instruments », dans *Gouverner par les instruments*, *op. cit.* p. 19.

¹⁸⁵ Néanmoins comme le souligne W. Parsons l'approche de Lasswell est sensiblement différente de la logique de l'EBP puisque l'entreprise du chercheur visait à politiser la connaissance en la rendant accessible aux citoyens alors que la logique de l'EBP vise à dépolitiser la production de connaissance et son utilisation « *The Lasswellian approach, above all, is about ensuring that knowledge is politicised, whereas for EBPM, the aim is quite the opposite: to de-politicise and managerialise knowledge production and its utilization* » Wayne PARSONS, « From Muddling Through to Muddling Up - Evidence Based Policy Making and the Modernisation of British Government », (2002) *Public Policy and Administration* p. 43-60.

¹⁸⁶ Jerry LEE, « Is Evidence-Based Government Possible? », 2004, p.5.

de synthèses de ces connaissances afin de les rendre accessibles aux personnes décisionnaires ou aux praticiens de ces secteurs. Ce mouvement cherche ainsi à influencer la production de connaissances dans les milieux scientifiques s'intéressant à l'action publique afin que ses résultats soient davantage utilisables par les décideurs publics, mais aussi à développer un nouveau système de management de l'apprentissage issu de données scientifiques au sein des institutions politiques¹⁸⁷.

Figure 1: La double logique de l'*Evidence-Based Policy*



L'EBP se caractérise par l'explicitation et la codification de ces pratiques. Cette démarche se présente donc comme un instrument de traduction entre les travaux scientifiques et le champ pratique auquel ils s'appliquent. Le choix des méthodes scientifiques à favoriser doit prendre en compte l'existence de travaux réalisés dans un domaine d'application, mais aussi leur force de persuasion auprès des praticiens susceptibles d'utiliser les résultats qu'elles produisent. Comme nous l'avons déjà évoqué précédemment, les travaux d'H. Marks sur le développement de *evidence-based medicine* aux États-Unis démontrent que la méthode expérimentale par assignation aléatoire était devenue la méthodologie privilégiée de cette pratique grâce, d'une part, à un travail de promotion effectué par les statisticiens souhaitant conduire des études avec cette méthode, mais aussi à sa simplicité de démonstration pour les praticiens médicaux non statisticiens¹⁸⁸.

Cependant, comme l'avaient souligné les travaux de R. A. Fisher et de D. Campbell, la validité interne d'une expérimentation aléatoire n'est pas suffisante pour préconiser la généralisation d'un « traitement ». Le passage à la prise de décision nécessite de fait l'identification d'un assez grand nombre d'études sur des traitements similaires dans différents contextes et de procéder à leur synthèse. Cette opération est appelée revue systématique (*systematic review*) et peut être définie comme l'utilisation de méthodes explicites et transparentes afin de procéder à la synthèse de résultats d'évaluations existantes et de projets de recherche produits dans le but de clarifier si une intervention donnée est efficace¹⁸⁹.

Lors de la stabilisation de la pratique « *evidence-based* » en santé publique, dans les années 1980 et 1990, est alors élaboré un protocole standard de synthèse de travaux. Cette méthodologie a été formalisée par la Collaboration Cochrane fondée en 1993 en Grande-Bretagne. Cette organisation, à but non lucratif, est un réseau indépendant de praticiens de santé et de chercheurs réunissant à l'heure actuelle des membres de 120

¹⁸⁷ Wayne PARSONS, *op. cit.* p. 4

¹⁸⁸ Harry MARKS, *op. cit.*

¹⁸⁹ Hansen RIEPER, « The evidence movement, the development and consequences of methodologies in review practices », (2009) 15 *Evaluation* p.141-163.

pays¹⁹⁰. Ce réseau organise la conduite par ses membres de revues systématiques dans différents domaines de santé publique selon un protocole stabilisé en 1994 dans le « Cochrane Handbook for systematic reviews »¹⁹¹. Ce protocole de synthèse propose des grilles d'évaluations de la qualité des études produites et place la méthode expérimentale en haut de cette hiérarchie méthodologique. Le protocole que doivent suivre ses membres dans leur conduite de revues systématiques, décrit dans son manuel, est le suivant :

Encadré 3 : Protocole des revues systématiques proposé par la Collaboration Cochrane

Le protocole de revues systématiques proposé par la Collaboration Cochrane¹⁹² 1. Définir une question de recherche <i>Sur le modèle : [type d'intervention] pour résoudre [quel problème] pour quel [type de public] .</i> 2. Définir une stratégie de recherche <i>Examiner le maximum d'études, Limiter les biais de sélection d'études, Formaliser la stratégie de recherche d'études et en garder trace afin qu'elle soit reproductible, Consulter les experts.</i> 3. Définir des critères incluant et excluant <i>Délimiter précisément le type d'intervention, de public, sur quelle durée, avec quels outils de comparaison (protocole expérimental avec placebo ou témoin), et l'impact étudié.</i> 4. Appréciation de la qualité des études <i>Etablir un protocole fixe de critères communs. La méthode expérimentale par assignation aléatoire en double aveugle est l'étalon or, suivi ensuite des études quasi expérimentales.</i> 5. Synthétiser les résultats <i>Si assez d'études expérimentales ont été identifiées, procéder à une méta-analyse, c'est-à-dire au calcul statistique de l'effet moyen de l'intervention étudié. La méta-analyse se présente sous la forme d'un graphique en arbre (forest plot) représenté sur le logo de cette organisation, il calcule l'effet moyen de l'intervention en prenant en compte les intervalles de confiance de chaque étude.</i>	
	

Ce modèle d'instrument *d'evidence-based* repose donc sur l'existence d'un nombre suffisant d'études conduites par la méthode expérimentale par assignation aléatoire afin de nourrir ses bases de données internationales et de procéder à des méta-analyses. En effet, jusqu'en 2005, les revues systématiques conduites par la collaboration Cochrane concluent, dans plus de 50% des cas, qu'ils ne disposent pas d'un nombre suffisant de travaux rigoureux pour déterminer l'efficacité du dispositif¹⁹³. Elle associe, de fait, le développement des pratiques « basées sur la preuve » avec celui de la méthode expérimentale et promeut donc le développement de cette méthode scientifique dans le secteur auquel elle s'applique.

Ainsi *l'evidence-based policy* peut faire référence d'une part à un instrument très précis d'appréciation de travaux scientifiques et de leur synthèse et d'autre part à une volonté politique de développer l'institutionnalisation du recours à des travaux scientifiques dont la méthodologie reste à définir. Comme tout instrument transféré d'une organisation à une autre, les modalités d'incarnation de *l'evidence-based policy* reflètent les contextes institutionnels et les traditions méthodologiques des organisations la mettant en place¹⁹⁴. Elle témoigne aussi de l'équilibre des forces au sein de la communauté des scientifiques ou des experts conduisant des travaux sur l'évaluation des

¹⁹⁰ Source: <http://www.cochrane.org/about-us/history>

¹⁹¹ JPT HIGGINS et S GREEN (dir.), *Cochrane Handbook for Systematic Reviews of Interventions*, The Cochrane Collaboration, 2011.

¹⁹² JPT HIGGINS et S GREEN *op. cit.*

¹⁹³ Hansen RIEPER, *op. cit.*

¹⁹⁴ Hansen RIEPER, *op. cit.*

interventions publiques et ayant un accès privilégié aux acteurs de l'action publique. De fait, le lien de *evidence-based policy* avec la méthode expérimentale par assignation aléatoire dépend de l'influence des promoteurs du modèle standardisé dans les différents contextes et de la présence de propositions alternatives de hiérarchie méthodologique. Une comparaison de l'incarnation de *evidence-based policy* en Grande-Bretagne, aux États-Unis et dans les instances internationales du développement, nous permet d'observer deux modèles distincts, l'un où la pratique expérimentale est encore peu répandue et proposant ainsi de forger ses propres outils en prenant en compte une diversité de travaux existants, et l'autre où la culture et la pratique expérimentale sont plus répandues et dont les promoteurs de l'EBP s'appuient exclusivement sur la conduite de la méthode aléatoire.

1.2.1.2. Le modèle original britannique, l'EBP par des réformes structurelles de l'État et une diversité de preuves

Le gouvernement britannique est le premier à théoriser, légitimer et mettre en œuvre l'EBP. Dès son arrivée au pouvoir en 1997, le gouvernement du *new labour* dont le slogan de campagne est « *What matters is what works* » (ce qui compte c'est ce qui marche), s'emploie à mettre en place ce nouveau mode de gouvernement¹⁹⁵. David Blunkett, alors secrétaire d'État à l'emploi et l'éducation déclare dans un discours intitulé « *Influence or Irrelevance: Can Social Science Improve Government?* » : « *La science sociale devrait être au cœur de l'élaboration des politiques publiques (policy making). Nous avons besoin d'une révolution dans les relations entre le gouvernement et la communauté de recherche en sciences sociales - nous avons besoin des chercheurs en sciences sociales pour nous aider à déterminer ce qui marche et pourquoi, et quelles sortes d'initiatives politiques ont des chances d'être les plus efficaces* »¹⁹⁶. Un autre rapport de 1999, produit par le Cabinet et intitulé « *Professional Policy Making for the Twenty-First Century* » résume cette démarche : « *l'action publique doit s'appuyer entièrement [soundly] sur la preuve de ce qui marche* » et « *les instances du gouvernement doivent améliorer leur capacité à utiliser les preuves* »¹⁹⁷.

Cette nouvelle équipe gouvernementale est caractérisée par sa rupture avec la tradition idéologique travailliste et sa volonté d'afficher un nouveau mode de gouvernement guidé par le pragmatisme, se dégageant du « gouvernement d'opinion »¹⁹⁸. Cette volonté s'est incarnée dans l'organisation, en 1999, de deux conférences nationales sur *evidence-based policy* par l'Association de centres de recherche en sciences sociales (ARCISS) et le département de politique publique de l'université de Londres en collaboration avec le centre Cochrane d'Oxford ouvrant sur la création du Centre de Ressource pour *evidence-based policy* (*Resource Centre for Evidence-Based Policy*)¹⁹⁹. Cette initiative a été institutionnalisée au sein du gouvernement britannique par le Livre blanc sur la modernisation du gouvernement publié en mars 1999 qui définit l'EBP comme « [l'aide] à la prise de décision bien informée sur des politiques publiques, des

¹⁹⁵ William SOLESBURY, *Evidence Based Policy: Whence it Came and Where it's Going*, ESRC UK Centre for Evidence Based Policy and Practice, Queen Mary, University of London, octobre 2001, p. 7

¹⁹⁶ Traduction de l'auteure à partir de l'anglais « *Social science should be at the heart of policy making. We need a revolution in relations between government and the social research community – we need social scientists to help to determine what works and why, and what types of policy initiatives are likely to be most effective.* » David BLUNKETT, *Blunkett rejects anti-intellectualism and welcomes sound ideas*, DfEE News, 2000

¹⁹⁷ Traduction de l'auteure à partir de l'anglais : « *policy making must be soundly based on evidence of what works* » (...) « *government departments must improve their capacity to make use of evidence* » CABINET OFFICE, *Professional Policy Making for the Twenty-First Century*, London, Cabinet Office, 1999, p. 40.

¹⁹⁸ Jerry LEE, « *Is Evidence-Based Government Possible?* », 2004, 1.

¹⁹⁹ Ron AMANN, « *Foreword* », dans Sandra M NUTLEY, Peter C SMITH et H. T. O DAVIES, *What works? : evidence-based policy and practice in public services*, Bristol, The Policy Press, 2000.

programmes et des projets en plaçant la meilleure preuve possible issue de la recherche au cœur du développement des politiques publiques et de leur mise en œuvre »²⁰⁰. S'en est suivi la création de plusieurs unités d'aide à la décision rattachées au Cabinet comme l'Unité de la performance et de l'innovation (*Performance and Innovation Unit*) ainsi que le Centre d'étude pour la gestion et les politiques publiques (*centre for management and policy studies*), actif dès 1999, auquel il incombait de nourrir la conception de politiques publiques par de récents travaux scientifiques existants²⁰¹.

Si *l'evidence-based policy* ne suggère pas que l'action publique soit littéralement *basée* sur la preuve, elle appelle cependant à une délimitation officielle de ce qui peut être considéré comme une donnée probante, suffisamment fiable, pour éclairer la décision. En Grande-Bretagne, est considérée comme preuve fiable une diversité de travaux issus de différentes méthodes et disciplines scientifiques. Dans le Livre blanc sur la modernisation de l'État, les preuves considérées comme légitimes sont : « *Le savoir des experts, les travaux de recherche publiés, les recherches existantes, la consultation des parties prenantes, les évaluations de politiques publiques existantes, internet, les résultats de consultations, l'appréciation financière d'options politiques, et ceux des modélisations économiques et statistiques* »²⁰². Des observateurs internes de l'État britannique précisent cependant que dans cette liste extensive, les recherches scientifiques, les statistiques, les évaluations, les modèles économiques et la connaissance d'experts sont les plus utilisés²⁰³. La difficulté réside alors dans l'identification des études existantes sur les sujets d'intérêt pour l'État, le jugement de leur rigueur et leur synthèse. Cette ouverture à une variété de preuve distingue l'EBP de la médecine des preuves (*evidence-based medicine*) car cette dernière a quant à elle établi une hiérarchie des preuves selon la méthode utilisée dans la démonstration scientifique et place la méthode d'évaluation par assignation aléatoire au sommet de cette hiérarchie²⁰⁴. Les promoteurs britanniques de la politique basée sur la preuve ne sont pas dans l'ignorance de cette hiérarchie et placent eux aussi cette méthode parmi les plus rigoureuses, tout en délimitant son périmètre de validité et en préférant le produit d'autres méthodes mixtes ou qualitatives lorsqu'ils cherchent à comprendre les mécanismes à l'œuvre dans l'intervention publique²⁰⁵. Un rapport gouvernemental de 2000, intitulé « *Adding it up* », identifie les tâches de l'État afin de favoriser l'EBP²⁰⁶. L'amélioration de la qualité des études produites est mentionnée, mais le travail fondamental est surtout identifié autour de la restructuration de l'action publique et de ses institutions pour favoriser une culture de dialogue avec les travaux scientifiques²⁰⁷. La seconde dimension de *l'evidence-based policy* tendant vers une standardisation de la production scientifique sur l'action publique selon les standards de la méthode expérimentale est donc très peu présente dans le modèle Britannique.

²⁰⁰ P DAVIES, « What is Evidence-Based Education? », (1999) 47-2 *British Journal of Educational Studies* pp108-121.

²⁰¹ Ron AMANN, *op. cit.*

²⁰² Traduction de l'auteure à partir de l'anglais : "Expert knowledge; published research, existing research; stakeholder consultations; previous policy evaluations; the Internet; outcomes from consultations; costings of policy options; output from economic and statistical modeling" CABINET OFFICE, *Modernising Government White Paper*, London, Center for Management and Policy Studies, 1999, p. 33.

²⁰³ Sandra M NUTLEY, Peter C SMITH et H. T. O DAVIES, « Introducing evidence-based policy and practice in public services », dans *What works? : evidence-based policy and practice in public services*, Bristol, The Policy Press, 2000.

²⁰⁴ Agnès LABROUSSE, « Nouvelle économie du développement et essais cliniques randomisés: une mise en perspective d'un outil de preuve et de gouvernement », (2010) 7 *Revue de la régulation*.

²⁰⁵ Ron AMANN, *op. cit.*; Sandra M. NUTLEY, H. T. O. DAVIES et Nick TILLEY, *op. cit.*

²⁰⁶ CABINET OFFICE, *Adding it up, improving analysis and modelling in Central Government*, London, Performance and Innovation Unit, 2000.

²⁰⁷ En témoigne notamment ce passage : "This study has found, however, that demand for good analysis is not fully integrated in the culture of central Government." P.12 CABINET OFFICE, *Adding it up, improving analysis and modelling in Central Government*, *op. cit.* p. 5.

Une nouvelle production de données évaluatives par la conduite de projets innovants rigoureusement évalués est aussi privilégiée dans l'incarnation britannique de l'EBP. Le rapport du Cabinet du gouvernement britannique intitulé « *Trying it out* » suggérait un recours souhaité systématique à cette démarche en ces termes: « *l'introduction à grande échelle de nouvelles interventions publiques (...) devrait dans la mesure du possible être précédée par des programmes pilotes suivis de près (closely monitored)* »²⁰⁸. Néanmoins, la méthode par assignation aléatoire était présentée comme une méthode d'évaluation rigoureuse, mais dont la légitimité était restreinte à un type particulier de questionnements et de type d'interventions.

D'autres instruments de l'EBP sont institutionnalisés par ce gouvernement, comme notamment les analyses d'impact (*impact assessments*). Ces analyses *ex ante* sont même devenues obligatoires pour toute intervention publique engendrant une variation de coûts, de nouvelles obligations d'information, de fardeau administratif, de redistribution, ou concernant des directives européennes²⁰⁹. Leur but est « *d'aider les décideurs publics à réfléchir aux raisons de l'intervention du gouvernement, de peser les options variées pour atteindre un objectif et de comprendre les conséquences d'une intervention proposée* »²¹⁰.

En parallèle à ces initiatives étatiques, des initiatives non gouvernementales sont organisées afin de promouvoir la réalisation d'un modèle particulier d'*evidence-based policy* correspondant aux standards de l'*evidence-based medicine*. Des membres de la Collaboration Cochrane, que nous avons mentionnée précédemment ont, à partir de 1998, commencé à réunir des chercheurs et des acteurs de l'action publique pour discuter de la diffusion de ce modèle de pratiques aux secteurs de l'action sociale. Ils décident de fonder en 2000 la Collaboration Campbell, dont le nom rend hommage au théoricien de la « société expérimentale », comme institution sœur pour conduire des revues systématiques dans le domaine de l'action sociale, de l'éducation et de la criminologie sur le même modèle que les revues systématiques de la collaboration Cochrane²¹¹. Faisant le constat que la plupart des travaux expérimentaux étaient conduits aux États-Unis et non en Grande-Bretagne, ils décidèrent de prendre contact avec R. Boruch, expérimentateur américain renommé, et de lui demander de mobiliser un réseau international d'acteurs favorables à ce projet²¹². C'est par cette première initiative que le courant de l'*evidence-based policy* a commencé à se diffuser au sein d'autres organisations via la promotion d'un modèle standardisé s'appuyant sur la conduite d'expérimentation aléatoire et leur synthèse.

La rhétorique de l'EBP s'est maintenue après le changement de gouvernement et jusqu'à aujourd'hui, néanmoins, il semble qu'elle s'est peu à peu orientée principalement vers la production de preuves et moins vers l'institutionnalisation de leur prise en compte. En témoigne ce passage du discours du Ministre de l'Éducation M. Gove en 2010 : « *En effet, je veux voir plus de données générées par la profession afin de montrer ce qui marche, des informations plus claires sur les techniques d'enseignement qui donnent des*

²⁰⁸ Roger JOWELL, *Trying it out : the role of « pilots » in policy-making : report of a review of government pilots*, Great Britain. Cabinet Office. Strategy Unit, London, Cabinet Office, Strategy Unit, 2003, p. 7.

²⁰⁹ Philip DAVIES, « The State of Evidence-Based Policy Evaluation and its Role in Policy Formation », (2012) 219 *National Institute Economic Review* 41-52, p.46.

²¹⁰ Traduction de l'auteure à partir de l'anglais : « *to help policymakers to fully think through the reasons for government intervention, to weigh up various options for achieving an objective and to understand the consequences of a proposed intervention* » de BIS (2011), *Impact Assessment Guidance When to do an Impact Assessment*, London, Department of Business, Innovation and Skills, p.4 cité dans Philip DAVIES, « The State of Evidence-Based Policy Evaluation and its Role in Policy Formation », (2012) *National Institute Economic Review*, p.46.

²¹¹ http://www.campbellcollaboration.org/resources/research/Methods_Links.php

²¹² Anthony PETROSINO, « Reflections on the Genesis of the Campbell Collaboration », (2013) 8 *The Experimental Criminologist* 9- 12.

résultats, des recherches plus rigoureuses et scientifiquement robustes sur les pédagogies qui réussissent, et de véritables évaluations indépendantes d'interventions mises en œuvre. Nous avons besoin davantage d'action publique s'appuyant sur des données probantes [Evidence-based policy making], et pour que cela marche, nous avons besoin de plus de preuves. »²¹³

Ce courant s'est développé au début des années 2000 en Australie, où son institutionnalisation a été moins importante et surtout concentrée sur les interventions publiques dans le domaine de la santé²¹⁴. Il s'est diffusé aussi en Nouvelle-Zélande dans le domaine de l'éducation²¹⁵. C'est principalement au modèle américain auquel nous souhaitons nous intéresser ici puisque ce dernier a particulièrement positionné la méthode expérimentale comme instrument de l'*evidence-based policy*, proposant une nouvelle déclinaison de la « société expérimentale » promue 40 ans plus tôt par Donald Campbell. La particularité du modèle américain vis-à-vis du modèle anglais est de faire reposer l'EBP principalement sur la production d'évaluations plus rigoureuses ayant recours à la méthode expérimentale par assignation aléatoire.

1.2.1.3. L'incarnation américaine, l'EBP par l'expérimentation aléatoire, un retour à la « société expérimentale »

La promotion de cette nouvelle démarche de conception de l'action publique a été promue aux États-Unis par l'organisation « à but non lucratif et non partisan » Coalition for Evidence-Based Policy, fondée en 2001²¹⁶. Cette organisation s'est donnée pour mission de « *promouvoir la conception de politiques publiques gouvernementales basées sur la démonstration rigoureuse de l'effectivité d'un programme* »²¹⁷. Elle défend une déclinaison de l'EBP s'appuyant principalement sur la conduite d'expérimentations évaluées par la méthode expérimentale par assignation aléatoire. On observe une filiation claire avec la promotion de la société expérimentale de la période étudiée précédemment puisqu'on compte dans son comité consultatif (*advisory board*) au moins deux acteurs de cette période, à savoir J. Gueron, ancienne présidente du MDRC (organisation qui comme nous l'avons vu a pris le relais du financement et de la conduite des expérimentations dans les années 1980), et R. Boruch, auteur de *Randomized Experiments for Planning and Evaluations* et qui a conduit l'entreprise de la création de la collaboration Campbell présentée précédemment. Cette coalition soutient explicitement la méthode expérimentale par assignation aléatoire comme l'outil

²¹³ Traduction de l'auteure à partir de l'anglais : "Indeed I want to see more data generated by the profession to show what works, clearer information about teaching techniques that get results, more rigorous, scientifically-robust research about pedagogies which succeed and proper independent evaluations of interventions which have run their course. We need more evidence-based policy making, and for that to work we need more evidence." Cité dans Philip DAVIES, « The State of Evidence-Based Policy Evaluation and its Role in Policy Formation », *op. cit.* p. 42

²¹⁴ Lisel O'DWYER, *A critical review of evidence-based policy making*, 58, Australian Housing and Urban Research, 2004; Greg MARSTON et Rob WATTS, « Tampering with the Evidence: A Critic Appraisal of Evidence-Based Policy-Making », (2003) 3-3 *The Drawing Board: An Australian Review of Public Affairs* 143- 163; Gary BANKS, *Challenges of Evidence - Based Policy - Making*, Australian Government, Productivity Commission, Australian Public Service Commission, 2009.

²¹⁵ Pour le lecteur intéressé par ce cas voir Adrienne ALTON-LEE, *Iterative Best Evidence Synthesis: Strengthening Research, Policy and Practice Links to Improve Outcomes*, 4th Annual Policy Conference: Policy Evolution, 29 mars 2006; Adrienne ALTON-LEE, *A Collaborative Knowledge Building Strategy to Improve Educational Policy and Practice: Work-in-Progress in the Ministry of Education's Iterative Best Evidence Synthesis Programme*, New Zealand Association for Research in Education National Conference 2004, Turning the Kaleidoscope, 25 novembre 2004.

²¹⁶ <http://coalition4evidence.org/staff/>

²¹⁷ Traduction de l'auteure à partir de l'anglais : "promote government policy-making based on rigorous evidence of program effectiveness"

de l'EBP, la positionnant comme « étalon or » des méthodes d'évaluation. Leur hiérarchie méthodologique est exprimée en ces termes ²¹⁸:

« 1. Une expérimentation aléatoire bien conçue, quand elle est faisable, est généralement la méthode la plus rigoureuse pour évaluer l'efficacité d'une intervention (...) »

2. Des groupes de comparaison bien appariés peuvent être un deuxième choix alternatif quand l'expérimentation aléatoire n'est pas faisable (...) »

3. Les autres méthodes plus communes, y compris les études « avant/après » et les groupes de comparaison appariés avec précaution peuvent être utiles pour générer des hypothèses sur ce qui marche, mais produisent souvent des conclusions erronées » ²¹⁹

Une enquête de B. Martin, composée d'interviews auprès des membres de cette coalition nous renseigne sur son entreprise de diffusion de ce modèle auprès des agences fédérales ²²⁰. Trois agences ont particulièrement été la cible de ce travail de Lobby, the Office of Management and Budget (OMB), le ministère de l'Éducation, et celui de la Justice. Son impact a été significatif sur l'OMB qui a organisé un comité avec les membres de la « Coalition » pour aboutir à la modification du guide de son outil d'évaluation des programmes publics (*Program Assessment Rating Tool*), transversal à toutes les agences du gouvernement et l'outil principal d'allocation de budget. Ce dernier précise maintenant que « les expérimentations par assignation aléatoire sont généralement les évaluations de la plus haute qualité et comportant le moins de biais pour démontrer l'impact d'un programme » ²²¹.

Cette coalition agit aussi par le biais de production de rapports prescriptifs comme par exemple, le rapport de 2002 adressé au ministère de l'éducation, dans lequel elle propose d'allouer des points additionnels aux évaluateurs incluant des protocoles expérimentaux par assignation aléatoire, dans son processus de sélection d'évaluateurs de projets ²²². La même année, l'Institut des sciences de l'éducation (*Institute of Education Sciences*), du même ministère, a créé le « What Works Clearinghouse » recensant les preuves scientifiques des programmes efficaces en éducation suivant le même modèle que la collaboration Cochrane ²²³. En 2003, le ministère de l'Éducation a publié deux documents donnant la priorité aux projets ayant un plan d'évaluation recourant à l'expérimentation aléatoire.

²¹⁸ COALITION FOR EVIDENCE-BASED POLICY, *Hierarchy of Study Designs for Evaluating the Effectiveness of STEM Education Project or Practice*, 2007.

²¹⁹ Traduction de l'auteur à partir de l'anglais : "A well-designed randomized controlled trial, where feasible, is generally the strongest study design for evaluating an intervention's effectiveness (...) II. Well matched comparison-group studies can be a second-best alternative when a randomized controlled trial is not feasible.(...)III. Other common study designs including pre-post studies, and comparison-group studies without careful matching can be useful in generating hypotheses about what works, but often produce erroneous conclusions."

²²⁰ Bernard H MARTIN, *The coalition for evidence-based policy: its impact on policy and practice*, William T. Grant Foundation, 2004.

²²¹ Traduction de l'auteur à partir de l'anglais : "randomized controlled trials are generally the highest quality, unbiased evaluations to demonstrate the actual impact of a program" Bernard H MARTIN, *The coalition for evidence-based policy: its impact on policy and practice*, op. cit. p. 2.

²²² Cette incitation fut effective puisque 30% des candidats évaluateurs ont proposé cette méthode d'évaluation COALITION FOR EVIDENCE-BASED POLICY, *Bringing Evidence-Driven Progress To Education: A recommended Strategy for the U.S. Department of Education*, 2002.

²²³ Pour une description détaillée du rôle de cette organisation voir Russell GERSTEN et John HITCHCOCK, « What is Credible Evidence in Education? The Role of What Works Clearinghouse in Informing the Process », dans Christina A CHRISTIE, Melvin M MARK et Stewart I DONALDSON (dir.), *What counts as credible evidence in applied research and evaluation practice?*, Los Angeles, SAGE, 2009, p. 78-96.

**Notice du 4 novembre 2003 du ministère de l'Éducation
« Scientifically Based Evaluation Methods »**

“The Secretary proposes a priority for program projects proposing an evaluation plan that is based on rigorous scientifically based research methods to assess the effectiveness of a particular intervention. The Secretary intends that this priority will allow program participants and the Department to determine whether the project produces meaningful effects on student achievement or teacher performance. Evaluation methods using an experimental design are best for determining project effectiveness. Thus, the project should use an experimental design under which participants—e.g., students, teachers, classrooms, or schools—are randomly assigned to participate in the project activities being evaluated or to a control group that does not participate in the project activities being evaluated.

If random assignment is not feasible, the project may use a quasi-experimental design with carefully matched comparison conditions...

In cases where random assignment is not possible and an extended series of observations of the outcome of interest precedes and follows the introduction of a new program or practice, regression discontinuity designs may be employed...

Proposed evaluation strategies that use neither experimental designs with random assignment nor quasi-experimental designs using a matched comparison group nor regression discontinuity designs will not be considered responsive to the priority when sufficient numbers of participants are available to support these designs.” ...²²⁴

“Identifying and Implementing Educational Practices Supported by Rigorous Evidence: A User Friendly Guide”, produit par le ministère de l'Éducation en décembre 2003

“To establish “strong” evidence, one needs “randomized controlled trials that are well-designed and implemented” plus “trials showing effectiveness in two or more typical school settings, including a setting similar to that of your schools/classrooms”

- To establish “possible” evidence, one needs “randomized controlled trials whose quality/quantity are good but fall short of ‘strong’ evidence; and/or comparison-group studies . . . in which the intervention and comparison groups are very closely matched in academic achievement, demographics, and other characteristics”*

- Types of studies that do not comprise even “possible” evidence include pre-post studies, non closely matched comparison group studies, and meta-analyses that include the results of lower quality studies.*

- If, according to these criteria, one does not have either “strong” nor “possible” evidence, then, “one may conclude that the intervention is not supported by meaningful evidence”.²²⁵*

Ces initiatives ont été relayées par le “Committee on research in education of the national research council” qui a organisé en septembre 2003 un atelier intitulé « *Implementing randomized field trials in Education* ». Ces liens du Ministère de l'Éducation avec la « coalition » se sont aussi concrétisés par la nomination de Jon Baron, président de cette organisation, au Bureau national pour les sciences de l'éducation (*National Board for Education Sciences*) de 2004 à 2011, dont il a pris la présidence en 2011²²⁶.

On retrouve la même stratégie, en 2003, avec le ministère de la Justice. La « coalition » a publié un rapport prescriptif qui a permis la création d'un site web répertoriant les programmes efficaces dans ce domaine²²⁷. Néanmoins, selon les participants à l'enquête

²²⁴ U.S. Department of Education. (2003a, November 4). Scientifically based evaluation methods. Federal Register, 68(213), 62445–62447. RIN 1890-ZA00. (p. 62446)

²²⁵ U.S. Department of Education, Institute of Education Sciences, National Center for Education Evaluation and Regional Assistance. (2003, November). Identifying and implementing educational practices supported by rigorous evidence: A user friendly guide

²²⁶ Source: <http://coalition4evidence.org/staff/>

²²⁷ COALITION FOR EVIDENCE-BASED POLICY, *Bringing Evidence-Driven Progress To Crime and Substance-Abuse Policy: A Recommended Federal Strategy*, 2003.

de B. Martin, le travail de persuasion de cette coalition n'a pas nécessité beaucoup d'efforts, ces agences étant acquises à ce type de démarche²²⁸. Cependant, l'insistance sur l'emploi de la méthode aléatoire est critiquée par plusieurs interviewés, ils estiment que dans beaucoup de types d'interventions, cette méthode n'est pas pertinente. Une seconde étude de 2009, a procédé à des entretiens avec les interlocuteurs de cette organisation dans les agences fédérales, elle souligne, à nouveau, cette insistance de l'organisation autour de l'utilisation des expérimentations aléatoires. L'auteur de l'enquête synthétise les critiques de plusieurs fonctionnaires, qui estiment que l'imposition de cette méthode comme "étalon or" est la principale faiblesse de cette organisation : « *A brief summary of their critiques is that, in their view, the Coalition overstates and oversimplifies the benefits of RCTs, thereby leading to the application of RCTs to inappropriate situations and ultimately hurting the cause of program evaluation. They would like to see the Coalition more fully address 1) the situations in which RCTs are inappropriate and 2) the value of non-RCT methodologies, both in these situations and more generally as a complement to RCTs. Some critics call their position "methodological pluralism".* »²²⁹. Une des recommandations principales de ce rapport est d'ailleurs d'abandonner la terminologie « d'étalon or », quand la Coalition se réfère à la méthode aléatoire, et de nuancer ses propos qui sont estimés comme allant à l'encontre de l'efficacité de la promotion de l'evidence-based policy.

On peut remarquer que ces principales initiatives ont trait uniquement au pendant de l'EBP portant sur la production de la connaissance et non sur le management de la prise en compte des connaissances par les instances de gouvernement. Cependant, dans certains domaines d'interventions fédérales comme la prévention de la violence, de la consommation de drogues, ou des grossesses adolescentes, des ministères ont adopté ce que C. Weiss appelle l'« utilisation imposée » (*imposed use*)²³⁰. Ce dispositif soumet les financements de projets, dans ces domaines d'intervention, à la démonstration scientifique au préalable de leur efficacité par une évaluation. Un panel d'experts est réuni pour étudier ces évaluations et leur rigueur afin d'établir une liste de projets performants pouvant faire l'objet de financement. L'étude de C. Weiss & alii sur ce dispositif dans le cadre de la prévention contre la consommation de drogue souligne les biais de cette démarche. D'une part, les experts participants peuvent avoir des conflits d'intérêts et favoriser les programmes auxquels ils ont participé dès la conception et, d'autre part, il semble que plus un programme est diffusé et évalué, plus la synthèse de ses évaluations tend vers une estimation de son impact proche de zéro, alors que des projets peu évalués seront davantage susceptibles d'entrer dans la liste des projets recommandés²³¹.

La poursuite de ces initiatives perdure sous le gouvernement de Barack Obama qui, dès son discours d'investiture, a déclaré au peuple américain que son administration s'appliquerait à se demander « *si [son gouvernement] marche, s'il aide des familles à trouver des emplois à un salaire décent, avec un système de santé accessible, et une retraite digne. La réponse sera oui, nous continuerons. La réponse sera non, le programme s'arrêtera.* »²³². P. R. Orszag, directeur du ministère du Budget (*Office of*

²²⁸ Bernard H MARTIN, *The coalition for evidence-based policy: its impact on policy and practice*, op. cit. p. 4.

²²⁹ Monica HERCK, *The coalition for evidence-based policy: its role in advancing evidence-based reform, 2004-2009*, William T. Grant Foundation, 2009, p. 9.

²³⁰ Carol H. WEISS, « The Fairy Godmother and Her Warts: Making the Dream of Evidence-Based Policy Come True », (2008) 29-29 *American Journal of Evaluation*, p.30.

²³¹ Carol H. WEISS, op. cit..

²³² Traduction de l'auteure à partir de l'anglais : "whether it works, whether it helps families find jobs at a decent wage, care they can afford, a retirement that is dignified. Where the answer is yes, we intend to move forward. Where the answer is no, programs will end" « President Barack Obama's Inaugural Address | The White House ».

Management and Budget) a soutenu cette ambition avec un communiqué de juin 2009 intitulé : « *Building Rigorous Evidence to Drive Policy* ». Par ce communiqué, il appelle à investir davantage dans des évaluations de qualité et présente la démarche d'« imposed used », évoquée plus haut pour certains types d'intervention, en ces termes : « *on insiste simplement sur le fait que la plupart de l'argent aille vers des programmes soutenus par la meilleure preuve disponible* »²³³. La meilleure preuve disponible doit être ici comprise comme celle produite par la méthode expérimentale, puisque nous avons vu préalablement que cette administration l'a positionnée à la tête de sa hiérarchie méthodologique²³⁴. La logique des différentes initiatives d'*evidence-based policy* concernant 6 domaines d'intervention est la suivante : sélectionner un problème social important, identifier des programmes mis en place pour résoudre ce problème dont l'efficacité est prouvée par des expérimentations aléatoires ou autres recherches rigoureuses, obtenir des fonds pour généraliser ces programmes et continuer à les évaluer²³⁵. En plus de ces initiatives, le budget présidentiel de 2011 a réservé des fonds pour l'évaluation des programmes - à titre d'exemple le ministère de l'emploi a reçu plus de 60 millions pour évaluer ses dispositifs -²³⁶.

Ainsi, le modèle américain de l'EBP est axé principalement sur la redéfinition des critères de scientificité des évaluations des projets financés par l'État fédéral autour de l'étalon or que constitue la méthode expérimentale par assignation aléatoire. Cette tendance est issue de la rencontre d'un contexte national déjà familiarisé avec cette méthode depuis le mouvement de la « société expérimentale » et d'un travail de cadrage cognitif d'un groupe de pression, la « Coalition », en imposant cette méthode comme instrument majeur de l'EBP. Les promoteurs de ce modèle estiment que le principal obstacle à l'EBP est imputable à l'absence de production d'évaluations d'impact rigoureuses pour renseigner l'action publique sur l'efficacité de ses interventions. Ce même cadrage cognitif de l'EBP est, à la même période, diffusé auprès des organisations internationales de développement. Le modèle standardisé de l'EBP se concentrant sur la qualité de la production de la preuve par la méthode expérimentale devient alors, au milieu des années 2000, le modèle dominant de l'EBP.

1.2.1.4. La diffusion dans les organisations internationales de développement, la production de preuves comme priorité de l'EBP

La promotion de l'*evidence-based policy* ne s'est pas limitée à des gouvernements nationaux. À partir du milieu des années 2000, cette nouvelle manière de conduire des interventions publiques a aussi été promue auprès des organisations internationales intervenant dans les politiques de développement. Dans les années 2000, le milieu du développement et particulièrement celui des économistes du développement est traversé par un débat sur l'efficacité de l'aide internationale. Ces années se caractérisent par la

²³³ Traduction de l'auteure à partir de l'anglais : "we only insist that most of the money go toward the programs backed by the best available evidence" Peter ORSZAG, « Building Rigorous Evidence to Drive Policy | The White House ».

²³⁴ Cette préférence est réaffirmée en 2009 : "With a team of powerful OMB officials fully committed to the value of experimental evaluations, the Obama Administration lost little time in launching its initiative to expand evidence-based social programs" Ron HASKINS et Jon BARON, *The Obama Administration's evidence-based social policy initiatives: An overview*, Coalition for Evidence-Based Policy.

²³⁵ La procédure est la suivante : 1. Select an important social problem that would make individual citizens and the nation better off if the problem could be successfully addressed by social policy. 2. Identify model programs addressed to the problem that have been shown by randomized trials or other rigorous research to significantly reduce the problem. 3. Obtain funds from Congress to scale up evidence-based programs of this type that attack the problem in accord with the verified models. 4. Make the funds available to government or private entities with a track record of good performance to replicate the successful model programs and to develop new model programs. 5. Continuously evaluate the projects as they are implemented to ensure they are faithfully implementing the model program and producing good results. Ron HASKINS et Jon BARON, *op. cit.*

²³⁶ Ron HASKINS et Jon BARON, *op. cit.*

reconnaissance par la Banque Mondiale et le PNUD (Programme des Nations Unies pour le Développement) de l'échec du consensus de Washington²³⁷. Ce consensus formulé par la Banque mondiale, le Fonds monétaire international et le Trésor public des États-Unis recommande d'appuyer la croissance des pays en développement sur des réformes néolibérales, par exemple, en réduisant l'intervention étatique, en équilibrant les budgets et en privatisant les entreprises publiques. Face à ce constat d'échec par les organisations internationales du développement, s'engage alors un débat parmi les économistes du développement à la fois sur les causes de cet échec, mais aussi sur les nouvelles modalités d'action à adopter pour favoriser le développement. D'une part, des économistes comme J. D. Sachs soutiennent que si les fonds étaient rendus disponibles la pauvreté pourrait être éradiquée, et d'autre part, d'autres économistes comme W. Easterly sont restés sceptiques sur les capacités de l'aide internationale²³⁸. Un autre questionnement traverse le débat sur la nature de l'aide à apporter. La question est de savoir si elle devait proposer des réformes de type « macro », par exemple la réforme de systèmes de santé, ou des interventions de type « micro » comme la distribution de moustiquaires²³⁹. À cette même période s'effectue aussi un déplacement de l'objectif premier de ces organisations vers la lutte contre la pauvreté plutôt que vers un développement plus général²⁴⁰.

Ce climat de doute sur l'efficacité ainsi que sur les modes opératoires des politiques de développement s'est accompagné d'une remise en cause du management de la connaissance par ses institutions et notamment de leur capacité à apprendre de leurs succès et de leurs erreurs, ainsi que de leurs systèmes d'information sur l'effectivité de leurs actions. La critique se concentre notamment, sur le peu d'études d'impact de « qualité » financées par ces organisations. Donnant naissance en 2003, au Abdul Latif Jameel Poverty Action Lab (J-PAL), fondé par E. Duflo, Abhijit Banerjee et Sendhil Mullainathan, au sein du département d'économie du Massachusetts Institute of Technology. D'abord financé par des fonds de start-up de l'université, ce réseau bénéficie, dès 2005, de larges donations d'un ancien élève du MIT, Mohammed Abdul Latif Jameel²⁴¹. Cette organisation est un « réseau mondial de professeurs unis par leur utilisation de l'évaluation aléatoire comme moyen de répondre à des questions critiques pour la lutte contre la pauvreté.(...) La mission de J-PAL est de réduire la pauvreté en faisant en sorte que les politiques soient fondées sur des preuves scientifiques »²⁴². Les membres de ce laboratoire, déjà influents dans le milieu du développement, proposent alors la conduite d'expérimentations aléatoires comme instrument de *evidence-based policy* dans le milieu du développement et comme solution pour agir malgré les controverses du moment. Un éditorial du *Lancet*, une revue médicale internationale, intitulé "The World Bank is finally embracing science" cite ces propos d'E. Duflo : "Creating a culture in which rigorous randomized evaluations are promoted, encouraged, and financed has the potential to revolutionize social policy during the 21st century, just as randomized trials revolutionized medicine during the 20th"²⁴³. On retrouve ce même appel dans l'ouvrage d'A. Banerjee « Making Aid Work », où il

²³⁷ Agnès LABROUSSE, « Nouvelle économie du développement et essais cliniques randomisés: une mise en perspective d'un outil de preuve et de gouvernement », (2010) 7 *Revue de la régulation*, 4.

²³⁸ Voir Sachs, Jeffrey D. 2005. *The End of Poverty: Economic Possibilities for Our Time*. New York: Penguin. Et Easterly, William. 2006. *The White Man's Burden: Why the West's Efforts to Aid the Rest Have Done So Much Ill and So Little Good*. New York: Penguin Press

²³⁹ Angus DEATON, « Instruments, Randomization, and Learning about Development », (2010) 48 *Journal of Economic Literature* 424- 455, 425.

²⁴⁰ Agnès LABROUSSE, *op. cit.* p. 4.

²⁴¹ <http://newsoffice.mit.edu/2005/gift>

²⁴² <http://www.povertyactionlab.org/fr/j-pal>

²⁴³ « The World Bank is finally embracing science, Editorial », (2004) *The Lancet*.

accuse la Banque Mondiale de « résistance à la connaissance » et propose la méthode expérimentale comme solution²⁴⁴.

Ce nouveau cadrage cognitif du problème des institutions du développement est relayé par un groupe de travail, « The Evaluation Gap Working Group », réuni à partir de 2004, par le Center for Global Development²⁴⁵. Ce groupe de réflexion compte notamment parmi ses membres J. Gueron du MDRC, par ailleurs membre de la « coalition for evidence based policy » que nous avons évoquée plus haut, E. Duflo, et F. Bourguignon, à l'époque économiste en chef et vice-président sénior de l'économie du développement à la Banque Mondiale. Ce groupe finalise son travail en 2006 par la publication d'un rapport destiné aux organisations internationales intitulé « *When will we ever learn ? Improving lives through impact evaluation* »²⁴⁶. Ce rapport plaide pour l'élaboration d'un nouveau système de connaissance au sein des organisations internationales, s'appuyant principalement sur la conduite d'évaluations d'impact rigoureuses, qui sont considérées comme la « pièce manquante » à leur apprentissage²⁴⁷. La méthode d'impact à privilégier est la méthode expérimentale par assignation aléatoire, en témoigne les exemples de bonnes pratiques présentés, qui ont en majorité eu recours à cette méthode²⁴⁸. Ce rapport appelle à la formation d'une organisation du type « Cochrane » qui, suivant les standards de *evidence-based practice*, mènerait des revues systématiques d'évaluations d'impact dans le domaine du développement et participerait au financement de ces expérimentations, que les membres de ce groupe considèrent comme des biens publics mondiaux. Ce rapport appelle à un agenda commun entre les pays financeurs et les organisations internationales pour garantir la conduite d'évaluation d'impact de qualité²⁴⁹. Cette recommandation verra le jour par la création de l'organisation non gouvernementale Initiative Internationale pour les Évaluations d'Impact (3IE), en mars 2008. Cette organisation finance des évaluations d'impact et en produit des revues systématiques sur le modèle de la collaboration Cochrane ou Campbell.

Ce travail de promotion de l'expérimentation aléatoire est devenu extrêmement efficace comme en témoigne l'augmentation du nombre d'expérimentations aléatoires. Il est d'ailleurs relayé aussi par des financeurs nationaux exigeant de la Banque Mondiale, un développement de la pratique d'évaluations. De plus, les États-Unis, sous les conseils de la « Coalition », ont requis que certains projets de la Banque Mondiale soient évalués rigoureusement avec une préférence pour les expérimentations aléatoires, de même, l'Espagne a alloué un fonds uniquement dédié à cet effet²⁵⁰. Le développement de la conduite d'évaluations d'impact a été aussi encouragé par le DIME (*Development Impact Evaluation Initiative*) créé en 2005²⁵¹. L'impact de ces initiatives s'avère conséquent puisque alors que de 2000 à 2004 seulement 20% des évaluations d'impact

²⁴⁴ Angus DEATON, *op. cit.* p. 348.

²⁴⁵ Ce centre fut fondé en 2001 par un ancien haut fonctionnaire américain Edward Scott Jr., C. Fred Bergsten, directeur du Peterson Institute for International Economics and Nancy Birdsall, qui auparavant était à la tête du département recherche de la Banque Mondiale source : <http://www.cgdev.org/page/about-cgd>. Ce centre est fondé par la foundation Bill et Melinda Gates et William et Flora Hewlett.

²⁴⁶ EVALUATION GAP WORKING GROUP, *When Will We Ever Learn? Improving Lives Through Impact Evaluation*, Center for Global Development, 2006.

²⁴⁷ *Ibid.* p. 2.

²⁴⁸ On peut aussi lire de manière plus explicite: "clinical trials of medications have become a standard and integral part of medical care. No physician would consider prescribing strong medications whose impact and potential side-effects have not been properly evaluated.(...) Where this method can be applied, it ensures that impact measurements are not confounded by systematic differences between beneficiary and control groups." *Ibid.* p.15.

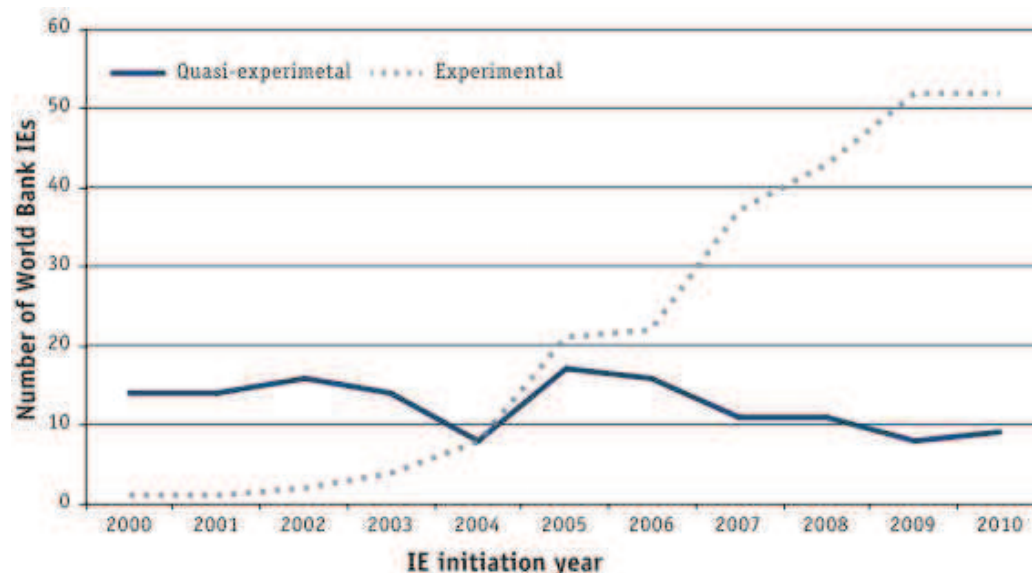
²⁴⁹ EVALUATION GAP WORKING GROUP, *op. cit.* p. 9.

²⁵⁰ Angus DEATON, *op. cit.* p. 438; Monica HERCK, *op. cit.*

²⁵¹ INDEPENDANT EVALUATION GROUP, *World Bank Group Impact Evaluations: Relevance and Effectiveness*, Washington, D.C, World Bank, 2012.

initiées par la banque mondiale ont eu recours à la méthode expérimentale par assignation aléatoire, entre 2009 et 2010, 83% l'ont utilisée²⁵².

Graphique 3 : Evolution du nombre d'évaluations d'impact à la Banque Mondiale utilisant la méthode expérimentale par assignation aléatoire (experimental) et quasi-expérimentale initiées entre 2000 à 2010 (source ²⁵³)



Cette nouvelle importance donnée à l'impact des programmes d'interventions plutôt qu'aux objectifs de développement constitue un changement radical dans la logique de l'intervention des organisations internationales et nationales de développement²⁵⁴. L'enjeu n'est plus de savoir si les indicateurs de développement révèlent une progression des pays bénéficiaires, mais si l'évolution de ces indicateurs est bien imputable aux programmes financés par ces organisations. Cette nouvelle orientation, relayée par la déclaration de Paris en 2005 consacrée à l'efficacité de l'aide au développement, insiste sur la mise en place de dispositifs permettant de renseigner sur l'impact de l'aide²⁵⁵. Cette démarche amène plusieurs agences nationales de développement à s'aligner sur ces nouveaux objectifs et à opérer une réforme de leur système d'évaluation donnant la priorité à la conduite d'évaluations d'impact « rigoureuses ». On note, en particulier une réforme des instances d'évaluation de l'Agence Française de Développement (AFD) à partir de 2006 qui intègre son unité d'évaluation à la division de la recherche afin de marquer cette nouvelle orientation des évaluations vers la production de connaissances correspondant aux standards de la production scientifique²⁵⁶.

Ainsi, l'étude de la diffusion du courant de *Evidence-Based Policy* témoigne d'une évolution de ce courant dans le sens d'une standardisation le rapprochant de la norme des pratiques *Evidence-Based* du secteur de la santé. Sa promotion aux États-Unis et au sein des organisations internationales, par des praticiens de l'expérimentation aléatoire tend à concentrer cet agenda de réformes sur la production et le financement d'évaluations d'impact de qualité, à savoir d'expérimentations aléatoires, au détriment

²⁵² Ibid. p. 31.

²⁵³ Ibid. p. 30.

²⁵⁴ Michael CLEMENS, *Applying evaluation to development and development aid*, Paris, Malaise dans l'évaluation: quelles leçons tirer de l'expérience du développement? 9ème conférence AFD/EUDN, 26 mars 2012.

²⁵⁵ Jean David NAUDET et Jocelyne DELARUE, *Fostering Impact Evaluations at Agence Française de Développement: A Process of In-house Appropriation and Capacity-Building*, Network of Networks on Impact Evaluation, janvier 2008.

²⁵⁶ Nous reviendrons sur la conduite d'expérimentations aléatoires par l'AFD dans notre étude du cas français Jean David NAUDET et Jocelyne DELARUE, *Fostering Impact Evaluations at Agence Française de Développement: A Process of In-house Appropriation and Capacity-Building*, Network of Networks on Impact Evaluation, janvier 2008.

de la modification des systèmes de prise en compte par les organisations publiques des données probantes. L'hypothèse portée par cette incarnation de ce courant est donc, que le principal obstacle à l'*Evidence-Based Policy* se situe du côté de la production de preuves et qu'une amélioration de la qualité des travaux mis à disposition des décideurs et financeurs publics augmente mécaniquement leur prise en compte par ces derniers²⁵⁷. Cette présentation de l'expérimentation aléatoire comme principal instrument de l'EBP devient alors la source d'une nouvelle période de débats, au sein de la communauté des chercheurs produisant des travaux d'évaluation de politique publique. Ils concernent autant le statut scientifique de cette méthode que son application à l'action publique et la vision de l'action publique qu'elle véhicule.

1.2.2. L'expérimentation aléatoire en débat : analyse des théories portées par l'expérimentation aléatoire comme instrument de l'EBP

La diffusion de l'*evidence-based policy* selon le modèle décrit précédemment n'a pas rencontré un consensus dans la communauté des chercheurs ou praticiens engagés dans des travaux d'évaluation de programmes ou de politiques publiques. La place particulière de l'expérimentation aléatoire, à la fois méthode de démonstration scientifique et d'évaluation, ainsi que ses différents secteurs d'intervention ont donné une dimension mondiale à ce débat à partir du milieu des années 2000. Au delà de l'intérêt pour ce débat en tant que tel pour rendre compte de la diversité des démarches scientifiques et professionnelles en évaluation des politiques publiques, son étude permet de mettre à jour les différentes théories portées par l'identification de l'expérimentation par assignation aléatoire comme instrument de choix de l'*Evidence-Based Policy*. Deux principaux sujets de débat se détachent, liés aux deux aspects de l'EBP, c'est-à-dire sa vocation à influencer la production scientifique d'une part et favoriser la prise en compte des travaux scientifiques par les acteurs de l'action publique d'autre part.

Le premier est de nature épistémologique et critique l'imposition d'une hiérarchie méthodologique aux travaux d'évaluation en s'intéressant notamment à la conception de la causalité ainsi qu'aux modalités de sa démonstration portées par cette méthode. Au cœur de ce débat baptisé la « guerre causale » (*causal war*), on note la critique de la conception mono-causale de cette méthode, et sa défense de la démonstration de la causalité uniquement par la comparaison avec un scénario contrefactuel.

Le deuxième débat, concerne la conception de l'action publique véhiculée par les hypothèses du choix de cette méthode comme instrument de l'EBP, ainsi qu'à la capacité de cette méthode à apporter des enseignements pertinents aux décideurs publics. Ses hypothèses d'un fonctionnement mécanique de l'action publique sont jugées non pertinentes et l'utilité des informations qu'elle produit pour les acteurs politiques, centrées exclusivement sur l'impact, notamment par la démarche de « revues systématiques », est remise en cause.

²⁵⁷ Esther DUFLO, « Evaluer l'impact des programmes d'aide au développement: le rôle des évaluations par assignation aléatoire », (2005) 19-2 *Revue d'économie du développement* pp 185-226.

1.2.2.1. « *The causal war* » le débat sur la pluralité méthodologique de la démonstration causale

*“When discussions of the role of experimental designs in evaluation became increasingly public a few years ago, I thought, “Didn’t we already settle this?” Almost 25 years ago, I organized and moderated a debate at the 1981 joint meeting of the predecessor organizations of the American Evaluation Association (AEA): the Evaluation Network and the Evaluation Research Society. The title of the debate was “Should the federal government mandate the use of experimental methods in evaluation?” The participants were Milbrey McLaughlin of Stanford University and Bob Boruch who currently is principal investigator of the What Works Clearinghouse and co-chair of the international Campbell Collaboration Steering Group. At the time, I published a couple of articles on the desirability of experimental methods and on non-causal alternatives (Smith, 1980, 1981). But that is the nature of fundamental issues; they may be temporarily resolved, but they will resurface later, in a different form, requiring a new, hopefully more thoughtful and sophisticated, resolution. I once commented to a senior colleague, David Krathwohl, that I thought I discerned a particular trend happening in evaluation. He observed that, if you lived long enough, you realized there were no trends, only cycles. So, I have lived long enough to see this fundamental issue return to our attention—indeed, if it ever left”.*²⁵⁸

Ces considérations de N. Smith dans son ouvrage *Fundamental issues in evaluation* reflètent l’étonnement de nombreux évaluateurs ayant participé aux précédents débats évoqués précédemment, face à cette nouvelle promotion de la méthode expérimentale comme «étalon or». Cette nouvelle ingérence de l’État américain et des organisations internationales dans le choix des méthodes d’évaluation en déterminant quelles études sont estimées comme assez rigoureuses pour constituer une démonstration scientifique, a été la source d’une nouvelle période de débats intense dans la communauté des évaluateurs tout au long des années 2000 que nous allons aborder ci-dessous. Néanmoins, nous n’assistons pas à la répétition exacte du même débat. Comme nous l’avons vu précédemment, les expérimentations aléatoires ont évolué elles sont davantage ancrées dans la démonstration de théories scientifiques. De fait, les termes de ce nouveau débat se sont maintenant déplacés sur le terrain de la démonstration de la causalité.

L’élément déclencheur s’inscrit dans les premières initiatives du ministère de l’Éducation, sous le conseil des membres de la *Coalition for evidence-based policy*, d’établir une hiérarchie méthodologique pour l’évaluation des projets qu’elle finance. La publication de ses guides successifs, positionnant la méthode expérimentale par assignation aléatoire comme la méthode la plus rigoureuse provoquent la publication d’un communiqué de la société américaine d’évaluation prônant une diversité méthodologique. Ce débat revêt de nouveaux enjeux pour les évaluateurs vis-à-vis du précédent. Lors de la période de la « société expérimentale », les évaluateurs ayant participé à la conduite d’évaluations expérimentales ont proposé ces méthodes d’évaluation de leur plein gré dont ils ont plus tard identifié les limites. Dans le contexte de *l’evidence-based policy*, la méthode aléatoire a été imposée aux évaluateurs comme idéal méthodologique par le gouvernement fédéral qui interfère de fait directement dans leur décision professionnelle de choix méthodologiques²⁵⁹.

Encadré 5 : Communiqué de l’association américaine d’évaluation « Scientifically Based Evaluation Methods »

²⁵⁸ Nick L. SMITH et Paul R. BRANDON (dir.), *Fundamental issues in evaluation*, New York, Guilford Press, 2008, p. 5.

²⁵⁹ *Ibid.* p. 18.

**AEA Statement:
American Evaluation Association Response to U. S. Department of Education
Notice of Proposed Priority, Federal Register
RIN 1890-ZA00, November 4, 2003**

“Scientifically Based Evaluation Methods”

The American Evaluation Association applauds the effort to promote high quality in the U.S. Secretary of Education’s proposed priority for evaluating educational programs using scientifically based methods. We, too, have worked to encourage competent practice through our Guiding Principles for Evaluators (1994), Standards for Program Evaluation (1994), professional training, and annual conferences. However, we believe the proposed priority manifests fundamental misunderstandings about (1) the types of studies capable of determining causality, (2) the methods capable of achieving scientific rigor, and (3) the types of studies that support policy and program decisions. We would like to help avoid the political, ethical, and financial disaster that could well attend implementation of the proposed priority.

(1) Studies capable of determining causality. Randomized control group trials (RCTs) are not the only studies capable of generating understandings of causality. In medicine, causality has been conclusively shown in some instances without RCTs, for example, in linking smoking to lung cancer and infested [sic] rats to bubonic plague. The secretary’s proposal would elevate experimental over quasi-experimental, observational, single-subject, and other designs which are sometimes more feasible and equally valid. RCTs are not always best for determining causality and can be misleading. RCTs examine a limited number of isolated factors that are neither limited nor isolated in natural settings. The complex nature of causality and the multitude of actual influences on outcomes render RCTs less capable of discovering causality than designs sensitive to local culture and conditions and open to unanticipated causal factors. RCTs should sometimes be ruled out for reasons of ethics. For example, assigning experimental subjects to educationally inferior or medically unproven treatments, or denying control group subjects access to important instructional opportunities or critical medical intervention, is not ethically acceptable even when RCT results might be enlightening. Such studies would not be approved by Institutional Review Boards overseeing the protection of human subjects in accordance with federal statute. In some cases, data sources are insufficient for RCTs. Pilot, experimental, and exploratory education, health, and social programs are often small enough in scale to preclude use of RCTs as an evaluation methodology, however important it may be to examine causality prior to wider implementation.

(2) Methods capable of demonstrating scientific rigor. For at least a decade, evaluators publicly debated whether newer inquiry methods were sufficiently rigorous. This issue was settled long ago. Actual practice and many published examples demonstrate that alternative and mixed methods are rigorous and scientific. To discourage a repertoire of methods would force evaluators backward. We strongly disagree that the methodological “benefits of the proposed priority justify the costs.”

(3) Studies capable of supporting appropriate policy and program decisions. We also strongly disagree that “this regulatory action does not unduly interfere with State, local, and tribal governments in the exercise of their governmental functions.” As provision and support of programs are governmental functions so, too, is determining program effectiveness. Sound policy decisions benefit from data illustrating not only causality but also conditionality. Fettering evaluators with unnecessary and unreasonable constraints would deny information needed by policy-makers. While we agree with the intent of ensuring that federally sponsored programs be “evaluated using scientifically based research to determine the effectiveness of a project intervention,” we do not agree that “evaluation methods using an experimental design are best for determining project effectiveness.” We believe that the constraints in the proposed priority would deny use of other needed, proven, and scientifically credible evaluation methods, resulting in fruitless expenditures on some large contracts while leaving other public programs unevaluated entirely.

La réaction à ce communiqué ne se fait pas attendre et témoigne de la division des pratiques évaluatives à ce sujet. Des membres de la même Association américaine d’évaluation publient un contre-communiqué intitulé « *the not AEA statement* » précisant que le précédent ne reflète pas leur point de vue. Dans ce communiqué, ces dissidents soutiennent la hiérarchie établie par le ministère de l’Éducation et

l'identification de la méthode aléatoire comme standard méthodologique des évaluations basées sur des méthodes scientifiques²⁶⁰.

Ces échanges houleux font alors place à l'organisation d'un débat à l'université de Claremont en 2004 entre Mark Lipsey, soutenant « *the not AEA statement* » et Michael Scriven soutenant le « *AEA statement* » et la pluralité méthodologique. À l'issue de cette rencontre, les deux chercheurs s'accordèrent sur le fait que la méthode aléatoire reste la meilleure méthode disponible pour estimer l'impact d'une intervention, donc sa causalité. Néanmoins, M. Scriven souligne que cette méthode peut uniquement être appliquée à une très faible proportion de projets éducatifs et qu'il existe maintenant de très bonnes alternatives méthodologiques pour établir les effets d'un projet alors que M. Lipsey refuse de reconnaître la pertinence de ces méthodes alternatives dans la démonstration de la causalité²⁶¹. La suite de ce débat prend la forme de la rédaction d'un ouvrage présentant différentes démarches et méthodes d'évaluation rigoureuses publié par l'université organisatrice du débat et intitulé « *What counts as credible evidence in applied research and evaluation practice ?* »²⁶².

Cet ouvrage, ne met cependant pas fin à la controverse au sein de la communauté des évaluateurs universitaires américains. Les termes du débat s'affinèrent néanmoins vers une concentration sur les modalités de la démonstration de la causalité en sciences sociales. Les défenseurs de la méthode expérimentale par assignation aléatoire soutiennent que cette méthode est la seule à même d'apporter une démonstration de l'effet causal, c'est-à-dire de l'impact d'une intervention publique. Parmi ses opposants se détachent alors deux positions, l'une soutenant que d'autres démonstrations de la causalité sont possibles grâce à d'autres méthodes et surtout d'autres conceptions de la causalité, l'autre soutient que l'estimation de l'impact d'une intervention ne doit pas être la mission unique des évaluations de programme. Ce débat sur la causalité se prolonge dans d'autres publications faisant se rencontrer ces différentes conceptions de la causalité comme notamment celui de T. Cook et M. Scriven²⁶³. Selon les défenseurs de la méthode expérimentale, la causalité d'une intervention, son impact, ne peut être démontré que par l'isolation de ce facteur en comparant deux situations identiques en tout point si ce n'est que l'une (le groupe test) bénéficie de l'intervention tandis que l'autre (le groupe témoin) n'en bénéficie pas. Ceci est appelé une analyse d'attribution. L'enjeu méthodologique est alors d'atteindre la plus grande similarité entre les deux groupes de comparaisons afin que le scénario contrefactuel, la situation sans intervention, soit assimilable à ce qu'aurait été la situation du groupe test s'il n'avait pas reçu le « traitement ». L'impact de l'intervention est donc estimé en soustrayant à l'évolution du groupe test, l'évolution du groupe témoin. La différence issue de ce calcul est alors considérée comme pouvant être uniquement issue du « traitement ».

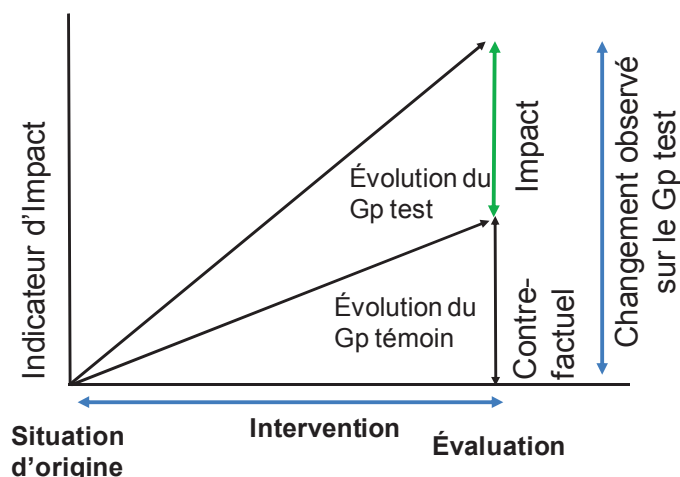
²⁶⁰ The Not AEA Statement: December 3, 2003

²⁶¹ Stewart I DONALDSON et Christina A CHRISTIE, « The 2004 Claremont Debate: Lipsey versus Scriven. Determining causality in program evaluation and applied research. Should experimental evidence be the gold standard? », (2005) 3 *Journal of Multidisciplinary Evaluation* 60-77, p.77.

²⁶² Stewart I DONALDSON, Christina A CHRISTIE et Melvin M MARK (dir.), *What counts as credible evidence in applied research and evaluation practice?*, Los Angeles, SAGE, 2009.

²⁶³ Thomas COOK et Michael SCRIVEN, « Contemporary Thinking About Causation in Evaluation: A dialogue With Tom Cook and Michael Scriven », (2010) 31-1 *American Journal of Evaluation* 105-117.

Figure 2: Logique de l'analyse d'attribution



Ce courant estime que la causalité peut être uniquement démontrée par la construction d'un scénario contrefactuel et la méthode expérimentale est alors considérée comme la meilleure puisqu'elle est la seule à pouvoir garantir une comparabilité des deux groupes en tous points grâce au tirage aléatoire. Or ce monopole de démonstration de la causalité a été discuté par un ensemble de chercheurs estimant que d'autres logiques de démonstration de la causalité existent et que d'autres méthodes sont alors pertinentes pour renseigner ces logiques²⁶⁴. Ces démarches alternatives ont été notamment présentées dans un rapport produit pour le DFID (*Department For International Development* du gouvernement britannique) en 2012, intitulé *Broadening the Range of Designs and Methods for Impact Evaluation*. La principale alternative est l'analyse de contribution qui propose d'établir une causalité sans procéder à l'élaboration d'un scénario contrefactuel²⁶⁵.

Cette démarche, s'intéressant à l'interaction d'une intervention publique et d'un contexte pour la production d'un effet, ne s'inscrit pas dans une conception monocausale et linéaire de la causalité. Elle s'inscrit dans la notion de *champ causal* et théorise le lien non pas entre un effet et une cause unique, mais entre un effet et un *faisceau causal*²⁶⁶ réunissant un ensemble de causes, dont chaque cause ne peut à elle seule provoquer l'effet final, mais dont la synergie avec les autres causes est susceptible de créer l'effet en question²⁶⁷. Par exemple, une allumette ou un grattoir séparément ne peuvent faire du feu, c'est la rencontre de ces deux éléments qui produit ce résultat. Néanmoins, ces deux causes ne sont pas nécessaires puisque on peut produire du feu par d'autres combinaisons d'éléments. L'évaluateur doit donc déterminer si l'intervention publique a agi comme un *mécanisme* activé et activant les autres éléments du faisceau causal pour atteindre l'effet souhaité²⁶⁸. Pour vérifier l'imputabilité du changement constaté à l'intervention publique, l'évaluateur procède à un découpage de l'intervention en plusieurs étapes (Activités, Résultats, Impacts intermédiaires, Impacts finaux) afin d'ouvrir la « boîte-noire » du programme. Puis, il identifie les facteurs de contexte, ou

²⁶⁴ Ces différentes logiques de démonstration de la causalité sont présentées dans le chapitre : Sue FUNNELL et Patricia ROGERS, « Causal Inference », dans *Purposeful program theory, effective use of theories of change and logic models*, Jossey-bass, 2011, p. 469-499.

²⁶⁵ John MAYNE, « Addressing Attribution Through Contribution Analysis: Using Performance Measures Sensibly », (2001) 16-1 *The Canadian Journal of Program Evaluation* 1- 24.

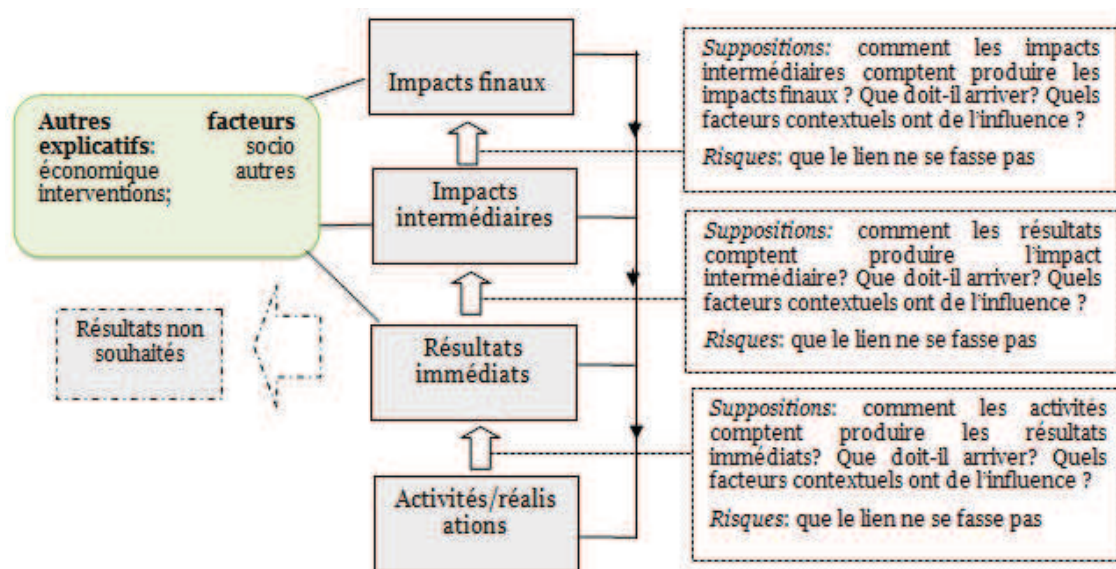
²⁶⁶ Merci à Jacques Toulemonde pour cette proposition de traduction du terme anglais original « causal package »

²⁶⁷ Barbara BEFANI, « Models of Causality and Causal Inference », dans *Broadening the Range of Designs and Methods for Impact Evaluation*, coll. DFID, 2012 p. 12.

²⁶⁸ Ray PAWSON, *Assessing the quality of evidence in evidence-based policy: why, how and when?*, ESRC Research Methods Programme, working paper.

causes extérieures, facilitant ou compliquant la production de l'effet, et leur interaction avec l'intervention publique. Il doit ensuite évincer les explications rivales c'est-à-dire vérifier que d'autres facteurs non liés à l'intervention n'ont pas été la cause principale du passage d'un maillon à l'autre de la chaîne. L'efficacité des *mécanismes* d'une intervention reste donc dépendante du contexte qui l'environne et peut favoriser ou freiner son enclenchement.

Figure 3: Logique de l'analyse de contribution



Comme le montre le diagramme ci-dessus, cette démarche identifie dans un premier temps la structure de l'intervention (activités/résultats/et impacts souhaités), elle étudie ensuite l'interaction de cette intervention avec son contexte de mise en œuvre en mettant à jour les suppositions et les risques liés à ce dernier pouvant activer ou empêcher l'activation du passage d'une étape à une autre. Elle cherche aussi à identifier d'autres facteurs explicatifs des impacts estimés afin de s'assurer qu'ils sont bien imputables au dispositif. Cette démarche, appelée aussi « explicative » s'inscrit dans le courant de l'évaluation basée sur la théorie, elle permet d'étudier la variabilité des effets d'un dispositif selon les contextes de mise en œuvre. S'appuyant sur l'épistémologie du réalisme critique, cette démarche ne considère pas qu'un programme marche ou ne marche pas en tant que tel, mais met à jour ses mécanismes et leur effectivité selon les contextes de leur mise en œuvre²⁶⁹.

1.2.2.2. Le débat sur l'utilité et la pertinence des résultats produits par la méthode expérimentale par assignation aléatoire pour l'action publique

Le deuxième aspect du débat se concentre sur l'utilité pour la programmation politique de l'information produite par la méthode expérimentale par assignation aléatoire ainsi qu'à l'influence de cette dernière sur la conception des programmes publics. Dans un premiers temps, nous présentons les arguments relevant la propension de cette hiérarchie méthodologique à favoriser le financement d'interventions d'un type particulier. Nous développons ensuite, les critiques de sa capacité à produire des enseignements généraux dépassant le contexte de la production des résultats expérimentaux, en abordant

²⁶⁹ Eric W. K. TSANG et Kai-Man KWAN, « Replication and Theory Development in Organizational Science: A Critical Realist Perspective », (1999) 24-4 *The Academy of Management Review* 759-780; Ray PAWSON et Nick TILLEY, *Realist Evaluation*, British Cabinet Office, 2004.

notamment les limites de l'estimation de l'impact « moyen ». Enfin, nous évoquons les critiques liées à son application dans le cadre des dynamiques de l'action publique.

Une première critique porte sur l'influence du monopole de scientificité alloué à cette méthode sur la nature des interventions publiques. Les défenseurs de ce modèle soutiennent qu'à l'image de la conduite des essais cliniques donnant leurs conclusions avant la mise sur le marché d'un médicament, l'efficacité d'une intervention publique devrait être établie par la méthode expérimentale avant sa généralisation²⁷⁰. Néanmoins, ils reconnaissent aussi que ce type d'évaluation ne peut s'appliquer qu'à un type particulier d'interventions publiques. Comme le soulignent les travaux de Rossi et Freeman, celles-ci doivent être simples - au sens d'inverse de complexe -, porter sur une intervention circonscrite et si possible standardisées avec un objectif précis²⁷¹. De fait, le couplage de cette méthode avec l'EBP tendrait vers le financement exclusif de projets de type « micro » et irait à l'encontre d'interventions « macro » davantage systémiques et complexes²⁷². Cette potentielle réorientation des interventions publiques a été notamment dénoncée dans le milieu du développement puisqu'elle implique des modalités d'interventions de lutte contre la pauvreté se concentrant sur des micro-interventions isolées et non sur des réformes structurelles sur la répartition des richesses, la lutte contre la corruption ou l'établissement de services publics de santé²⁷³. Cette hiérarchie méthodologique favorise donc le type d'interventions standardisées financées par des fondations privées ou des organisations internationales sans interférer avec la structure des gouvernements locaux. Ce danger de la diffusion du monopole de cette méthode et son influence sur la nature des projets est relevé par M. Ravallion. Cet économiste, devenu depuis 2012, économiste en chef et vice président sénior à la Banque Mondiale, souligne qu'aujourd'hui des chercheurs refusent d'évaluer des programmes si la randomisation n'est pas possible, que les doctorants cherchent continuellement des projets à randomiser et que les fondations deviennent de plus en plus réticentes à financer des évaluations non expérimentales²⁷⁴. La diffusion de cette hiérarchie pourrait donc contribuer à modifier considérablement la nature des interventions publiques dont le financement serait soumis à une évaluation positive par cette méthode.

La principale critique adressée à cette méthode et son utilisation politique porte surtout sur sa capacité à produire des informations pertinentes pour la décision publique. Est d'abord soulignée, l'implication de la perception mono-causale des interventions publiques proposée par cette méthode. Cette conception impacte la validité externe de ses résultats et notamment leur sensibilité aux contextes locaux. Comme le souligne A. Deaton, l'absence de prise en compte de la particularité de l'échantillon expérimental, et notamment de la cohérence du traitement étudié avec les autres politiques publiques en œuvre sur le même territoire ou les contextes culturels, est un obstacle pour en déterminer la généralisation. Comme les participants à l'expérimentation ne sont pas tirés au sort parmi la population à laquelle le traitement est susceptible d'être généralisé, elle peut présenter des caractéristiques particulières limitant la généralisation des résultats²⁷⁵. Pour soutenir son argument, il mobilise notamment les remarques du concepteur du programme Progressa, souvent mis en exergue comme un programme

²⁷⁰ Esther DUFLO, « Evaluer l'impact des programmes d'aide au développement: le rôle des évaluations par assignation aléatoire », *op. cit.*

²⁷¹ Peter H. ROSSI et Howard E. FREEMAN, *op. cit.* p. 290.

²⁷² Jos VAESSEN, « Challenges in impact evaluation of development interventions: opportunities and limitations for randomized experiments », *University of Antwerp* 2010, p.15.

²⁷³ Sanjay G. REDDY, *Randomise This! On Poor Economics*, 2 août 2013, p. 2.

²⁷⁴ Martin RAVALLION, « Should Randomistas rule? », *Economist's Voice* 2009.

²⁷⁵ Angus DEATON, *op. cit.* p. 445.

ayant fait ses preuves par la méthode expérimentale et dont la diffusion est à promouvoir. Ce programme allouait une somme d'argent aux familles sous réserve d'une scolarisation de leurs enfants ainsi que leur participation à des soins médicaux²⁷⁶. Santiago Levy, son concepteur souligne néanmoins que son exportation à d'autres pays doit prendre en compte les programmes sociaux déjà existants dans ces territoires afin de s'assurer qu'ils n'entrent pas en contradiction avec l'intervention et doit être aussi attentif aux capacités du système éducatif à accueillir davantage d'élèves²⁷⁷. En reprenant les travaux de N. Cartwright, A. Deaton relève un contraste entre l'attention accordée par cette méthode à la validité interne et sa démarche beaucoup moins rigoureuse dans ses préconisations d'extrapolation des résultats positifs à d'autres contextes²⁷⁸. Ce manque d'attention dans les pratiques *evidence-based* dans les processus de généralisation des résultats à partir de travaux expérimentaux a d'ailleurs été aussi souligné dans le cadre de l'*Evidence-based Medicine*²⁷⁹.

D. Rodrik illustre les problèmes liés à la généralisation en prenant l'exemple de l'expérimentation de Cohen et Dupas visant à trancher le débat des économistes pour savoir s'il fallait mieux distribuer les moustiquaires gratuitement ou faire payer un prix minimum pour en augmenter l'usage²⁸⁰. Le résultat de cette expérimentation conclut que la distribution gratuite est plus efficace et a été interprétée comme un argument permettant de clore ce débat international. Néanmoins, D. Rodrik souligne que de nombreuses contingences locales constituent des obstacles à cette généralisation et peuvent aussi contribuer à expliquer le succès de la gratuité dans ce contexte particulier²⁸¹. Il note que le territoire de l'expérimentation est particulier du fait des campagnes préalables qui avaient déjà diffusé auprès de la population les vertus de la moustiquaire ; aussi le public cible composé de femmes enceintes consultant les cliniques parentales était déjà volontaire à s'engager dans une démarche de santé et était bien informées sur les risques de la malaria ; enfin les moustiquaires étaient mises à la disposition des cliniques et donc l'expérimentation ne nous informe pas sur la potentialité d'organisation et de maintien d'un marché de l'offre de moustiquaire en cas de généralisation. Il en conclut que les résultats produits par cette expérimentation peuvent être uniquement considérés comme valides et rigoureux dans leur contexte, ce qui l'amène à souligner « *que la seule preuve « dure » [hard evidence] que les expérimentation aléatoires peuvent typiquement administrer tient à des questions qui sont d'une portée et d'une application si étroitement limitée qu'elles sont en elles-mêmes inintéressantes* »²⁸².

Une solution proposée par les promoteurs de la méthode expérimentale est alors d'appuyer la validité externe des expérimentations sur leur réplication. D'abord cette logique, en elle-même, semble difficilement tenable puisque, comme le souligne M. Ravallion, cela supposerait d'expérimenter toutes les modalités de mise en œuvre de programme ainsi que des options politiques de différents choix d'intervention à

²⁷⁶ Paul GERTLER, *The Impact of Progresa on Health*, International Food Policy Research Institute, 2000.

²⁷⁷ Angus DEATON, *op. cit.* p. 441.

²⁷⁸ Cartwright, Nancy. 2007a. "Are RCTs the Gold Standard?" *Biosocieties*, 2(1): 11–20. Citée par Angus DEATON, « Instruments, Randomization, and Learning about Development », (2010) 48 *Journal of Economic Literature* 424–455, p.450.

²⁷⁹ Voir un tableau recapitulative de ces arguments dans Dani RODRIK, *The new development economics: We shall experiment, but how shall we learn?*, Brookings Development Conference, 29 mai 2008, p. 22.

²⁸⁰ Cohen, Jessica, and Pascaline Dupas, "Free Distribution or Cost-Sharing? Evidence from a Randomized Malaria Prevention Experiment," *Global Economy and Development Working Paper 11*, The Brookings Institution, Dec 2007.

²⁸¹ Dani RODRIK, *The new development economics: We shall experiment, but how shall we learn?*, Brookings Development Conference, 29 mai 2008, p. 14.

²⁸² Traduction de l'auteure à partir de l'anglais : "the only truly hard evidence that randomized evaluations typically generate relates to questions that are so narrowly limited in scope and application that they are in themselves uninteresting." Dani RODRIK, *op. cit.* p. 5.

considérer²⁸³. La logique de réplcation et de synthèses proposée par les meta-analyses – calcul quantitatif de l’effet moyen d’un dispositif par la synthèse de plusieurs résultats d’expérimentations aléatoires –, semble déjà constituer un problème en soi. Baser l’estimation de l’efficacité d’un dispositif sur une moyenne de résultats ne permet pas de rendre compte des effets hétérogènes du traitement sur les différentes catégories de population ou selon les différents contextes auxquels il s’applique. Ainsi, avec cette méthode, un programme peut être jugé satisfaisant s’il a un impact négatif sur une partie de la population étudiée, mais que cet effet est compensé par un effet positif sur le reste de la population. Si certains chercheurs conduisant des expérimentations mènent des analyses *ex post* par sous-groupe afin de rendre compte de l’hétérogénéité des résultats, celles-ci ne bénéficient plus alors des standards de rigueurs contenus par la méthode aléatoire et doivent être alors sujettes aux mêmes critiques que les résultats produits par des méthodes non expérimentales²⁸⁴. Si l’on extrapole cette logique de moyenne au niveau de la méta-analyse, cela implique que des dispositifs peuvent être jugés comme efficaces alors qu’ils ont des impacts négatifs sur certaines populations ou dans certains contextes. D’ailleurs, comme le souligne C. Weiss, selon son expérience, plus un dispositif est expérimenté dans une variation de contextes, plus l’estimation de son impact moyen est susceptible d’être nul, du fait de la sensibilité de son efficacité à des contextes différents²⁸⁵.

Ainsi, une solution proposée par plusieurs auteurs est de compenser les failles explicatives de l’expérimentation et de ses modalités de synthèse en les combinant avec d’autres méthodes de collectes de données ou d’analyse permettant d’expliquer les variations de l’effet d’un programme selon les contingences de sa mise en œuvre et du contexte environnant. Dans les communautés des évaluateurs américains et des économistes du développement, des tentatives de réconciliation de démarches expérimentales et analytiques ont été entreprises par de nombreux travaux. Van der Knapp & Alii proposent d’articuler la méthode expérimentale avec une approche « réaliste » afin d’estimer l’impact tout en comprenant les mécanismes de succès ou d’échec²⁸⁶. H. White du 3iE (*international initiative for impact evaluation*) appelle aussi à un recours à une démarche évaluative basée sur la théorie afin d’expliquer des différences d’impact entre des mêmes dispositifs mis en œuvre dans des contextes différents et évalués par la méthode expérimentale²⁸⁷. Cet appel est aussi relayé par Joes Vaessen qui propose de coupler ces deux approches dès la conception de l’expérimentation²⁸⁸. Enfin, pour procéder à des synthèses de preuves afin d’éclairer des décisions publiques, R. Pawson suivant la logique de l’EBP britannique propose une méthode alternative, la « synthèse réaliste », prenant en compte une variété de travaux, y compris des expérimentations aléatoires, mais sans hiérarchie méthodologique, et axée sur l’identification de facteurs favorables ou défavorables à la mise en place de certains *mécanismes* présents dans plusieurs interventions publiques, comme par exemple l’incitation au changement de comportement par la gratuité d’un produit²⁸⁹.

²⁸³ Martin RAVALLION, « Should Randomistas rule? », *Economist’ Voice* 2009, p.18.

²⁸⁴ Angus DEATON, *op. cit.* p.439.

²⁸⁵ Carol H. WEISS, « The Fairy Godmother and Her Warts: Making the Dream of Evidence-Based Policy Come True ».

²⁸⁶ Leontien VAN DER KNAAP, Frans LEEUW, Stefan BOGAERTS et Laura NIJSSEN, « Combining Campbell Standards and the Realist Evaluation Approach, The Best of Two Worlds? », (2008) 29-48 *American Journal of Evaluation* 48- 57.

²⁸⁷ Howard WHITE, *Theory-Based Impact Evaluation: Principles and Practice*, International Initiative for Impact Evaluation, 2009.

²⁸⁸ Jos VAESSEN, *op. cit.* p. 24.

²⁸⁹ Ray PAWSON, *Evidence-based policy: a realist perspective*, London ; Thousand Oaks, Calif, SAGE, 2006.

Enfin, les critiques sur les risques de mauvaise mise en œuvre de cette méthode et de leur instrumentalisation par leur rencontre avec les acteurs de l'action publique sont deux critiques structurantes du débat lors de la période de « la société expérimentale » qui sont ici très marginales. Ceci est peut être dû au fait qu'elles aient déjà été traitées préalablement mais aussi probablement, suite au cadrage particulièrement scientifique de ce nouveau débat. Voyons la critique portant sur les bricolages de terrain nécessaires à la conduite des expérimentations pouvant endommager la validité interne des résultats. Cette dimension est évoquée mais encore peu appuyée sur des travaux empiriques. Elle fut d'abord soulignée par les praticiens de cette méthode, E. Duflo explique : « *Il peut donc y avoir conflit entre l'exigence de faire le projet de la manière la plus efficace possible et la moins chère possible du point de vue de l'organisation, et celle de respecter le protocole ; mais on arrive toujours à s'adapter. Les protocoles doivent être plus ou moins respectés, on est capable après sur le plan de l'analyse des données de prendre en compte des différences avec la situation idéale des laboratoires* »²⁹⁰. Cependant, ces ajustements sont peu mis en exergue dans la restitution des travaux et A. Labrousse appelle à rendre explicite les bricolages qu'implique cette technique, dont les incarnations peuvent varier selon les terrains auxquels elle s'applique²⁹¹. Selon les modalités des ajustements opérés, ceux-ci peuvent porter considérablement atteinte à la validité interne des résultats ainsi qu'à leur possibilité de généralisation. A. Deaton relève par exemple, des dérogations au tirage au sort dans l'expérimentation de M. Kremer et M. Edward testant l'efficacité du traitement des enfants dans les écoles sur leur assiduité, où ces dernières ont été choisies par ordre alphabétique²⁹². Les biais intervenant dans la collecte de données liées à l'interaction sociale entre les bénéficiaires ont aussi été soulignés par les travaux de J. Droitcour et M. Kovar²⁹³. Néanmoins, il semble que ces travaux s'intéressant aux conditions de mise en œuvre effectives ont été relégués, et font place à des débats évacuant la collecte de données primaires et s'intéressant principalement aux modalités d'analyse des données, leur interprétation et leur généralisation. Quant à la remise en question de la capacité des acteurs de l'action publique à s'engager dans une démarche scientifique, non partisane, et à ne pas instrumentaliser (*missuse*) les résultats issus de travaux de recherche, elle est quasiment absente du débat. Elle est évoquée par plusieurs travaux étudiant les effets de l'EBP en Grande-Bretagne, soulignant que beaucoup d'autres facteurs interviennent dans le modèle décisionnel politique et que les preuves sont surtout mobilisées quand elles vont dans le sens de l'agenda politique²⁹⁴. Aux Etats-Unis, C. Weiss souligne les intérêts en jeu dans les pratiques *evidence-based* qui peuvent être source de nombreux biais et qui ne sont pas uniquement guidées par la simple recherche du programme le plus efficace²⁹⁵. Nous souhaitons, par notre travail, éclairer d'abord la dimension de bricolage de l'expérimentation aléatoire en portant notre attention sur la dimension sociale de l'expérimentation, sur l'interaction qu'elle organise entre chercheurs, porteurs de projets et financeurs, et les compromis que cela implique. L'étude de cette relation nous renseigne aussi sur les intérêts politiques et scientifiques en jeu dans cette entreprise, ce qui nous permettra de comprendre dans quelle mesure la conduite

²⁹⁰ E. Duflo cite par A. DEATON, *op. cit.* p. 39.

²⁹¹ Agnès LABROUSSE, « Nouvelle économie du développement et essais cliniques randomisés : une mise en perspective d'un outil de preuve et de gouvernement », p. 22.

²⁹² Miguel, Edward, and Michael Kremer. 2004. "Worms: Identifying Impacts on Education and Health in the Presence of Treatment Externalities." *Econometrica*, 72(1): 159–217. Angus DEATON, *op. cit.*

²⁹³ Judith A. DROITCOUR et Mary Grace KOVAR, « Multiple Threats to the Validity of Randomized Studies », dans Nick L. SMITH et Paul R. BRANDON, *Fundamental issues in Evaluation*, The Guildford Press, New York, 2008, p. 65–88.

²⁹⁴ Sandra M. NUTLEY et Peter C. SMITH, « Evidence and the policy process », dans H. T. O. DAVIES, *What works? : evidence-based policy and practice in public services*, Bristol, The Policy Press, 2000; Jonathan S. DAVIES, « The Limits of Joined-up Government: Towards a Political Analysis », (2009) 87-1 *Public Administration* 80–96; Philip DAVIES, « The State of Evidence-Based Policy Evaluation and its Role in Policy Formation », (2012) 219 *National Institute Economic Review*.

²⁹⁵ Carol H. WEISS, « The Fairy Godmother and Her Warts: Making the Dream of Evidence-Based Policy Come True ».

d'expérimentations par assignation aléatoire s'inscrit dans une démarche *d'evidence-based policy*.

Au final, ces débats sur la rigueur méthodologique des évaluations ainsi que sur leur capacité à répondre aux besoins des décideurs publics ont contribué à engager un travail de formalisation et de théorisation de la rigueur des démarches alternatives à l'expérimentation aléatoire ainsi qu'une légitimation de leur pertinence vis-à-vis du contexte politique. Ce travail a abouti à une affirmation de l'identité professionnelle des chercheurs-évaluateurs²⁹⁶.

Ainsi, ce premier développement a permis de retracer la sociogenèse de la méthode expérimentale par assignation aléatoire et de *l'Evidence-Based Policy* et de démontrer que les développements de ces deux objets sont intimement liés.

Nous avons démontré que cette méthode scientifique n'est pas neutre, en particulier lorsqu'elle est utilisée dans le cadre de l'évaluation des politiques publiques. Elle porte en elle certaines théories mais ces dernières peuvent aussi varier selon le cadre d'utilisation et les acteurs la conduisant. L'étude de son utilisation au cours du XX^e siècle révèle un idéal constant contenu par cette méthode de contrôle de la subjectivité du chercheur par l'organisation d'une manipulation artificielle du phénomène observé. Cependant, les objectifs de son emploi dans le cadre de l'évaluation des politiques publiques ont varié selon les acteurs la conduisant. Son appropriation par les micro-économistes statisticiens, à partir du milieu des années 1990, en a fait un objet d'organisation de la rencontre entre le test de théories scientifiques et l'identification de l'efficacité de l'intervention publique. Cette utilisation couplée à *l'evidence-based policy* défend une seconde hypothèse consistant à imputer le manque de prise en compte des travaux scientifiques dans le cadre de l'action publique au manque de preuves rigoureuses existantes pour l'informer et estime que mécaniquement, le financement de la production de preuves augmentera leur impact sur la conception des politiques publiques.

Un premier enseignement est que la supériorité démonstrative de la méthode expérimentale par assignation aléatoire est issue de son inscription dans une entreprise sociale et politique, que ce statut est remis en question par une variété de chercheurs et de praticiens de l'évaluation, et qu'il peut être expliqué par un travail de lobby d'organisations ou d'individus praticiens de cette méthode, auprès des instances politiques. Un second enseignement, issu de l'étude du cas américain, est que l'expérimentation aléatoire semble prendre des formes variables selon les personnes la commanditant et la conduisant et selon le discours justificateur de son association avec l'action publique. Ces deux enseignements constitueront deux points d'attention guidant notre étude du cas français. Néanmoins, de prime abord, au regard de ce développement le cas français soulève plusieurs questionnements.

Nous avons vu que la promotion de la méthode expérimentale comme instrument majeur de *l'evidence-based policy* à un niveau national n'avait été possible que dans le contexte américain où cette méthode était déjà largement utilisée avant les années 2000. Or, la méthode expérimentale par assignation aléatoire n'a pas été utilisée pour l'évaluation des politiques publiques avant 2006 en France et a néanmoins été présentée comme la méthode phare de l'EBP par une instance de l'administration centrale française dès

²⁹⁶ Nick L. SMITH et Paul R. BRANDON *op. cit.* p. 19.

2009. Ce succès peut-il être uniquement imputé à un simple phénomène de mode et d'importation d'instruments ? Dans ce cas, comment expliquer que cette méthode n'ait pas été importée plus tôt au cours du XX^{ème} siècle en France puisqu'elle était utilisée aux Etats-Unis depuis cinquante ans ? Enfin, puisque nous avons vu que les modalités d'incarnation de l'expérimentation aléatoire et notamment sa propension à réorganiser l'interaction entre la science et l'action publique dépendait de l'identité de ses promoteurs et de leurs discours de légitimation, quels ont été les acteurs français de sa promotion et surtout quel modèle d'expérimentation ont-ils défendu ? Afin d'étudier ces éléments il est nécessaire de revenir d'abord en détail sur le cadre théorique ainsi que l'appareillage méthodologique que nous allons développer dans les deux chapitres suivants.

2. Etude d'un instrument d'action publique par une approche constructiviste de l'institution, positionnement théorique et implications méthodologiques

Le retour sur la sociogenèse de la méthode expérimentale par assignation aléatoire et son utilisation dans le cadre de *l'evidence-based policy* dans des contextes étrangers nous a offert un premier aperçu sur les modalités de son développement ainsi que son impact sur l'action publique. Les premiers constats développés ci-dessus nous ont guidée dans l'élaboration de cadres théoriques potentiellement pertinents pour appréhender de manière empirique ce phénomène sur le territoire français. Suivant l'agenda de recherche ouvert par l'étude des instruments d'action publique notre travail s'intéresse d'une part aux facteurs de l'émergence d'utilisation de cette méthode et d'autre part à ses effets, en l'étudiant comme une institution sociale.

Nous détaillons dans cette partie les deux appareillages théoriques mobilisés dans le chapitre 2 et le chapitre 3 de ce travail de recherche. Nous présentons dans un premier temps l'appareillage théorique et méthodologique mobilisé pour appréhender l'émergence d'une institution dans une conception constructiviste de cette dernière. Dans un second temps nous revenons sur l'appareillage théorique et méthodologique élaboré pour étudier des incarnations diverses d'une même institution d'apprentissage au gré du travail institutionnel des acteurs qui la composent, permettant de rendre compte de l'importance de la dimension interactionnelle dans cette entreprise.

2.1. L'Etude empirique de l'émergence d'une méthode d'évaluation de politiques publiques par la *théorie générale des champs* et le *travail institutionnel*

Cette partie expose les choix théoriques et empiriques du chapitre 2 de ce travail. La *théorie générale des champs* ainsi que le concept de *travail institutionnel* nous semblent l'alliage le plus pertinent pour appréhender une construction institutionnelle mouvante, issue d'une coalition d'acteurs venant de champs distincts. Ce travail de construction institutionnelle correspond à la structuration d'un *champ d'action stratégique* qui s'adapte au gré des résistances et des succès qu'il rencontre. Ce cadre théorique ouvre un programme de recherche sur les types de travaux institutionnels engagés par des groupes d'acteurs distincts, ainsi que l'importance des *pratiques* dans la structuration de ce processus.

2.1.1. Un cadrage théorique alliant la sociologie des champs et le néo-institutionnalisme sociologique.

La méthode expérimentale par assignation aléatoire a été l'objet de nombreuses publications scientifiques, évoquées précédemment, mais jusqu'à présent elle n'a pas été étudiée avec les outils de la sociologie de l'action publique²⁹⁷. L'identification d'un cadre théorique pertinent a donc nécessité la recherche d'outils théoriques et de concepts utilisés pour rendre compte de phénomènes proches, et éprouver leur pertinence pour cet objet.

Ce travail propose d'élaborer un cadre théorique permettant de mettre à jour des structures sous-jacentes et des mécanismes afin de rendre compte et d'expliquer l'émergence de cette méthode à une certaine période en France. L'étude de l'utilisation d'une méthode scientifique pourrait suggérer une inscription dans les travaux de la sociologie de l'expertise ou de la sociologie des sciences ; néanmoins, cet objet a la particularité d'organiser un recours au conseiller scientifique non pour son expertise sectorielle, mais pour sa maîtrise d'une méthodologie. En d'autres termes le scientifique est moins mobilisé pour son *savoir* que pour son *savoir-faire*. Nous nous sommes donc tournés davantage vers les cadres théoriques permettant de rendre compte de l'émergence de nouvelles techniques, d'innovations, et de leur diffusion, ainsi que l'étude de la quête de ressources pour les développer.

Dans son introduction, ce travail est déjà revenu sur la pertinence de comprendre la méthode expérimentale par assignation aléatoire comme un *Instrument d'Action Publique* (IAP)²⁹⁸. Cette conception dépasse la conception fonctionnaliste de l'instrument et ouvre un programme de recherche sur l'étude des luttes et des valeurs dont ils sont issus, puisqu'ils sont considérés comme des institutions sociales, dans l'acception large des néo-institutionnalistes²⁹⁹.

Comprenant alors cet objet comme une construction institutionnelle, nous choisissons des cadres théoriques qui appréhendent les coalitions d'individus orientés vers la formation d'une institution ou la promotion d'une cause. Dans un premier temps, nous justifierons notre choix du concept de *champ d'action stratégique* de N. Fligstein et D. McAdam, pour caractériser l'organisation collective ayant promu la méthode expérimentale et en expliquer les motivations³⁰⁰. Nous reviendrons sur les moteurs de l'émergence d'un tel champ en nuancant les conceptions posant *l'entrepreneur*

²⁹⁷ Les seuls travaux abordant cette méthode avec les outils de la sociologie des sciences ou de l'histoire se sont intéressés aux essais cliniques ayant recours à la méthode aléatoire. On note le travail de l'historien Harry Marks qui cherche à identifier les facteurs de développement de cette méthode en médecine aux États-Unis. On relève aussi les travaux de N. Dodier et J. Barbot sur la diffusion de cette méthode dans le cadre des traitements contre le Sida en France qui mobilise l'analyse de réseau. Ces deux travaux relèvent de la sociologie historique des sciences, mais n'ont pas la dimension politique de notre objet. Harry MARKS, *op. cit.*; Nicolas DODIER et Janine BARBOT, « Le temps des tensions épistémiques: Le développement des essais thérapeutiques dans le cadre du sida », (2000) 41-1 *Revue française de sociologie* 79-118.

²⁹⁸ Que l'on comprend suivant la définition de Patrick Le Galès et Pierre Lascombes, comme un « dispositif technique à vocation générique porteur d'une conception concrète du rapport politique/ société », Patrick Le Galès et Pierre Lascombes, « L'action publique saisie par ses instruments », *op. cit.* p. 14

²⁹⁹ Soit comme « un ensemble de règles et de procédures plus ou moins coordonnées qui gouvernent les interactions et comportements des acteurs et des organisations », traduction de l'auteure à partir de l'anglais « a more or less coordinated set of rules and procedures that governs the interactions and behaviors of actors and organizations » dans Powell, W., and P. Di Maggio, *The New Institutionalism in Organizational Analysis*. Chicago: University of Chicago Press, 1991

³⁰⁰ Neil FLIGSTEIN et Doug MCADAM, *A theory of fields*, New York, Oxford University Press, 2012.

institutionnel comme acteur unique maîtrisant un processus. Sans mettre de côté ce concept, nous précisons les limites d'une conception de toute puissance de ce dernier, en élargissant notre étude aux autres acteurs de ce processus, lesquels procèdent à un *travail institutionnel* de support ou de déstabilisation plus subtil³⁰¹. Enfin, nous expliquerons notre conception des modalités d'évolution de ce champ, en revenant sur la dimension chaotique de ce processus non linéaire, et le lien de dépendance de ce dernier avec les autres champs dans lequel il s'inscrit. Nous insisterons enfin sur la place centrale accordée au *travail institutionnel* cognitif dans ce phénomène.

2.1.1.1. L'Etude de l'émergence d'un groupement d'individus orienté vers la promotion d'une nouvelle organisation sociale

Ce travail s'inscrit dans la perspective théorique néo-institutionnaliste et sociologique dont les concepts permettent d'établir un langage commun entre la théorie des organisations, l'analyse de l'innovation ainsi que l'étude des mouvements sociaux. Le cadre théorique néo-institutionnel a été longtemps considéré comme un outil dédié à l'étude des phénomènes de stabilité institutionnelle³⁰². Néanmoins, depuis une trentaine d'années, un renouveau théorique de cette démarche propose des concepts pour analyser le changement institutionnel ainsi que les phénomènes d'émergence des institutions, en accordant notamment une nouvelle place à l'étude de l'agence des acteurs sur les institutions. Le développement de concepts comme le *champ organisationnel*, l'*entrepreneur institutionnel*, le *travail institutionnel* ou les *champs d'action stratégiques* s'inscrivent dans ce nouveau courant.

Depuis une trentaine d'années, l'agenda de recherche ainsi que les outils de l'approche théorique néo-institutionnelle ont été diversifiés et affinés afin de mieux appréhender l'émergence des institutions, notamment en considérant l'action des individus dans cette entreprise³⁰³. Dépassant l'étude statique des facteurs de la stabilité des institutions, ce courant du néo-institutionnalisme propose une approche dynamique de l'institution en élaborant des concepts pour appréhender ses processus de construction et d'émergence ainsi que, parfois, son déclin ou sa disparition. L'institutionnalisation est comprise comme un processus visant à objectiver les normes, en d'autres termes, « *un processus permanent qui tend à faire passer des règles, des savoirs, des dispositifs, des régularités et des rôles (...) pour nécessaires et fonctionnels, et qui peut même les soustraire à la discussion et les absolutiser* »³⁰⁴. S'intéresser à ce processus permet de mettre l'accent sur la finalité d'une action collective en soulignant les modalités d'organisation et de structuration de ce mouvement. Elle pose aussi la question de l'agence : qui a la capacité de s'extraire de ses propres routines institutionnelles pour forger une nouvelle entité ? Cette première partie propose d'aborder cette littérature et de revenir sur la conception de l'interaction de l'institution et de l'individu qu'elle propose.

³⁰¹ Thomas B. LAWRENCE et Roy SUDDABY, « Institutions and Institutional Work », dans Stewart CLEGG, C HARDY et T.B. LAWRENCE, *The Sage handbook of organization studies*, 2nd ed, London ; Thousand Oaks, Calif, Sage Publications, 2006, p. 215-254.

³⁰² Jacques LAGROYE et Michel OFFERLE, *Sociologie de l'institution*, Paris, Belin, 2010, p. 14.

³⁰³ Paul DiMAGGIO, « Interest and agency in institutional theory. », dans Lynne G. ZUCKER (dir.), *Institutional patterns and organizations: Culture and environment*, Cambridge, Mass, Ballinger, 1988, p. 3-22; Christine OLIVIER, « strategic responses to institutional processes », (1991) 16 *Academy of Management Review* 145- 179; Thomas B. LAWRENCE et Roy SUDDABY, « Institutions and Institutional Work », *op. cit.*

³⁰⁴ Jacques LAGROYE et Michel OFFERLE, *op. cit.* p. 13.

Une part importante des études portant sur l'émergence d'institutions s'inscrit dans l'héritage de la sociologie de l'innovation développée par M. Alkrich, M. Callon et B. Latour à la fin des années 1980³⁰⁵. Le néo-institutionnalisme sociologique fondé dans les années quatre-vingt, notamment par les travaux de W. Powell et P. DiMaggio, connaît un élan théorique croissant jusqu'à nos jours. Il permet d'expliquer les changements en contexte organisé, en ouvrant l'étude des processus de construction sociale des formes institutionnelles. Ce courant considère que le développement des organisations et institutions résulte de leur encastrement social et historique. Ces dynamiques institutionnelles, en grande partie cognitives canalisent les choix sociaux que font les organisations. Des travaux de la fin des années quatre-vingt et du début des années quatre-vingt-dix, émergent les concepts de *champ organisationnel* de W. Powell et P. DiMaggio et de *secteur organisationnel* de J. Meyer et R. Scott³⁰⁶. Ces concepts permettent de concevoir l'environnement comme un ordre social de niveau intermédiaire, construit socialement, et n'ayant pas la stabilité d'une institution, mais tendant à le devenir. Ils permettent de situer le lieu où se construisent et se diffusent les cadres cognitifs contribuant à la construction de l'institution.

Le concept de *champ organisationnel* désigne une agrégation d'organisations qui produisent des services ou des produits similaires. Ces organisations interagissent entre elles plus fréquemment qu'avec des acteurs extérieurs au champ, elles partagent un système de croyances et un sentiment d'appartenance, et interagissent autour d'un projet à construire ou d'une problématique à résoudre³⁰⁷. Ce cadre cognitif ouvre un nouveau domaine d'investigation sur l'identification des contours de ce champ, de ses acteurs, ainsi que de son discours unificateur. Le choix des organisations de s'inscrire dans ce champ résulte à la fois de représentations symboliques, de la diffusion de croyances, ainsi que d'un calcul stratégique, puisque les organisations estiment augmenter leurs chances de survie au sein de leur environnement institutionnel en rejoignant ce cercle. L'interaction entre ce cadre structurant d'une part et l'intentionnalité des acteurs d'autre part, à l'échelle de l'individu ou de l'organisation, constitue l'objet principal du renouveau néo-institutionnaliste.

L'utilisation du concept de champ pour appréhender l'émergence d'organisations ou d'institutions a fait l'objet d'un travail récent proposant une *théorie générale des champs*³⁰⁸. S'appuyant sur les travaux des néo-institutionnalistes, et de P. Bourdieu et L. Wacquant sur ce concept, N. Fligstein et D. McAdam proposent le concept de *champ d'action stratégique*. Ce dernier est défini comme un « *ordre social construit, de niveau intermédiaire, au sein duquel les acteurs (qui peuvent être individuels ou collectifs) ont*

³⁰⁵ Aussi bien les textes français sur les instruments d'action publique que les textes anglo-saxons sur l'entrepreneur institutionnel, le travail institutionnel font référence aux travaux de sociologie de l'innovation dont : Madeleine AKRICH, Michel CALLON et Bruno LATOUR, « A quoi tient le succès des innovations? 1: L'art de l'intéressement », (1988) 11-12 *Gérer et comprendre, Annales des Mines*; Madeleine AKRICH, Michel CALLON et Bruno LATOUR, « A quoi tient le succès des innovations? 1: L'art de l'intéressement », (1988) 11-12 *Gérer et comprendre, Annales des Mines*; Madeleine AKRICH, Michel CALLON et Bruno LATOUR, « A quoi tient le succès des innovations? 2: Le choix des porte-parole », (1988) 12 *Gérer et comprendre, Annales des Mines* 14-19; Michel CALLON, « Eléments pour une sociologie de la traduction, la domestication des coquilles Saint-Jacques et des marins-pêcheurs dans la baie de Saint-Brieuc », (1986) 36 *l'Année sociologique* pp 169-208.

³⁰⁶ P. DiMaggio et W. Powell, « The Iron Cage Revisited: Institutional Isomorphism and Collective Rationality in Organizational Fields », *op. cit.*; W.R. Scott et J.W. Meyer, « The organization of societal sectors; proposition and early evidence », in *The New Institutionalism in Organizational Analysis*, par W. Powell et P. DiMaggio (Chicago:University of Chicago Press, 1991), 109-142.

³⁰⁷ Cette définition croise celle de P. DiMaggio et W. Powell « key suppliers, resources and product consumers, regulatory agencies and other organization that produce similar services and products » et celle de Scott « the notion of field connotes the existence of a community of organization that partakes of a common meaning system and whose participants interact more frequently and fatefully with one another than with actors outside the field » Paul DiMAGGIO et Walter POWELL, *op. cit.*; W.R. SCOTT, *Institutions and organizations*, Thousand Oaks, California, Sage publication, 1995, p. 6.

³⁰⁸ Neil FLIGSTEIN et Doug MCADAM, « Toward a General Theory of Strategic Action Fields », (2011) 29-1 *Sociological Theory* 1-26, p.2.

conscience les uns des autres et interagissent sur la base d'une compréhension partagée (ce qui ne veut pas dire consensuelle) sur les buts du champ, les relations entre les personnes du champ (dont qui a le pouvoir et pourquoi), ainsi que les règles gouvernant l'action légitime dans le champ »³⁰⁹. Le champ est compris comme une unité d'action collective, une arène au sein de laquelle les acteurs sont en compétition pour des avantages³¹⁰. Il est composé de titulaires (*incumbents*), des acteurs qui ont une influence disproportionnée au sein du champ et dont les intérêts et vues se reflètent fortement dans l'organisation dominante du champ ; de challengers (*challengers*) qui détiennent des niches moins privilégiées, ils ont moins d'influence et parfois une vision alternative du champ, mais se conforment à l'ordre établi ; et enfin d'instances de gouvernance, qui vérifient la conformité aux règles du champ et peuvent être internes ou externes. Chaque champ se situe sur un continuum entre des activités de coopération consistant à construire une coalition politique pour maintenir le champ, et des activités de compétition entre plusieurs groupes au sein du champ³¹¹.

Ces différentes conceptions théoriques semblent assez proches, mais se distinguent par la manière dont elles conçoivent le processus d'émergence des champs. Poser la question de l'émergence suggère de s'intéresser au pouvoir d'agence des acteurs au sein des champs, et de déterminer quel type d'acteur détient ce pouvoir d'agir pour initier le changement. Il ouvre à d'autres questionnements : Comment expliquer qu'un individu, une organisation ou un ensemble d'individus s'engagent dans une activité concourant à institutionnaliser des pratiques, des normes, ou de véhiculer une symbolique commune ? Cette activité d'impulsion témoigne-t-elle d'une liberté de l'acteur vis-à-vis de cadres sociaux d'action, ou s'inscrit-elle dans une logique structurée par un environnement institutionnel préalable ? Le cas échéant, comment rendre compte d'une circularité d'emprise entre l'acteur et l'institution tout en étudiant le mouvement institutionnel ?

Les travaux de P. DiMaggio, S. Eisenstadt et C. Olivier placent l'entrepreneuriat institutionnel au centre du processus de constitution du champ³¹². Selon P. DiMaggio, « les nouvelles institutions émergent quand des acteurs organisés avec suffisamment de ressources voient en elles une opportunité de réaliser les intérêts auxquels ils accordent une valeur importante »³¹³. Cette perspective met l'accent sur le travail et l'initiative des acteurs pour influencer leur contexte institutionnel³¹⁴. Ces « acteurs organisés », ou *entrepreneurs institutionnels*, sont des personnes ou des organisations dont le rôle est de participer à la formulation et à la définition d'un problème commun, tout en proposant une solution pour le résoudre selon leurs propres enjeux et intérêts³¹⁵. Les *entrepreneurs institutionnels* sont perçus comme des acteurs à part, jouissant de ressources et de capacités exceptionnelles. Ils ont des aptitudes réflexives qui leur permettent de prendre

³⁰⁹ Traduction de l'auteure de l'anglais "A strategic action field is a constructed mesolevel social order in which actors (who can be individual or collective) are attuned to and interact with one another on the basis of shared (which is not to say consensual) understandings about the purposes of the field, relationships to others in the field (including who has power and why), and the rules governing legitimate action in the field" Neil FLIGSTEIN et Doug MCADAM, *A theory of fields*, op. cit. p. 9.

³¹⁰ Neil FLIGSTEIN et Doug MCADAM, *A theory of fields*, op. cit. p. 64.

³¹¹ *Ibid.* p. 15.

³¹² S. EISENSTADT, « Cultural orientations, institutional entrepreneurs, and social change: Comparative analysis of traditional civilizations », *American Journal of Sociology* 1980.85.840-869. Christine OLIVIER, « strategic responses to institutional processes », (1991) 16 *Academy of Management Review* 145-179.

³¹³ Traduction de l'auteure à partir de l'anglais "new institutions arise when organize actors with sufficient resources see in them an opportunity to realize interests that they value highly" Paul DiMAGGIO, « Interest and agency in institutional theory. », dans Lynne G. ZUCKER (dir.), *Institutional patterns and organizations: Culture and environment*, Cambridge, Mass, Ballinger, 1988, p. 3-22 p. 14.

³¹⁴ R GARUD, S JAIN et A KUMARASWAMY, « Institutional entrepreneurship in the sponsorship of common technological standards: The case of Sun Microsystems and Java. », (2002) 45-1 *Academy of Management Journal* 196-214.

³¹⁵ M. C. SUCHMAN, « Managing Legitimacy: Strategic and Institutional Approaches », (1995) 20 *Academy of Management Review* 571- 610; R DELBRIDGE et T EDWARDS, « Challenging conventions: Roles and processes during non-isomorphic institutional change », (2008) 61-3 *Human Relations* 299-325.

de la distance vis-à-vis de leur institution d'appartenance, et des capacités à mobiliser un soutien politique ainsi qu'à convaincre d'autres personnes d'adhérer au projet institutionnel³¹⁶. Ils s'apparentent à l'entrepreneur schumpetérien, médiateur entre les deux univers du marché et de l'innovation, et aux *acteurs réseaux* décrits par M. Callon dans la sociologie de la traduction. Ces entrepreneurs maîtrisent le processus de constitution du champ et *intéressent* d'autres acteurs à leur cause en en définissant les termes en leur faveur.

Comment expliquer que des acteurs soient plus aptes que d'autres à devenir entrepreneurs ? Les entrepreneurs sont davantage enclins à agir de par leur positionnement singulier dans l'espace social. Ce sont des *marginiaux sécants*, pour reprendre les termes de M. Crozier, ou des *médiateurs* qui sont à cheval entre deux mondes si l'on se rapporte aux travaux de P. Muller, à même de conduire des travaux de *généralistes* ou de *courtiers* selon la conception d'O. Nay et d'A. Smith. Cette appartenance plurielle leur permet d'entretenir une distance critique avec les construits sociaux des champs auxquels ils appartiennent. Les entrepreneurs configurent souvent les nouveaux champs pour mettre en lien les univers dans lesquels ils sont déjà positionnés.

Plusieurs auteurs nuancent aujourd'hui cette vision de toute puissance de l'entrepreneur dans ce processus d'émergence. On questionne notamment la maîtrise du processus de constitution de champ ou d'institution par l'entrepreneur, en prenant en compte le pouvoir d'agence d'autres acteurs du champ.

F. Bergeron propose le concept *d'entrepreneur frontière* qui tel *l'objet frontière*, concept forgé par S. L. Star et J. Griesemer en sociologie des sciences, est « *ouvert aux projections et aux manipulations des différents acteurs participants de ces univers* »³¹⁷. Cet entrepreneur ne cherche pas alors principalement à aligner les autres acteurs sur son programme, mais à rendre ambigu ce dernier ainsi que sa propre identité afin de rallier un maximum de support. Ainsi, *l'entrepreneur frontière* n'est pas juste un intermédiaire, qui transmet un message sans transformation ou un médiateur qui traduit selon M. Callon, B. Latour, bricole selon Castel et Friedberg, ou procède à un transcodage selon P. Lascoumes, des éléments mobilisateurs communs au champ. Il n'est pas face à un ensemble d'acteurs passifs, mais il est aussi « *agi* » par ces derniers. Cet entrepreneur peut « *relier des univers sociaux hétérogènes, sans prétendre modifier les logiques régulant chacun de ces univers, ni les identités des acteurs qui les composent* »³¹⁸. Il le fait en ajustant son discours, le contenu de l'entreprise du champ ou même la partie de son identité qu'il met en exergue selon les valeurs du groupe auquel il s'adresse. Une ambiguïté, polysémie, ou une absence de définition précise de l'entreprise est un des facteurs de succès du champ, puisqu'il permet de rallier un maximum d'intérêts divergents. Les « *entrepris* » sélectionnent les facettes de cette définition servant au mieux leurs intérêts et redéfinissent les objectifs du champ de manière localisée, en dehors du contrôle de l'entrepreneur, afin que ces objectifs servent leurs intérêts. Ces auteurs font l'hypothèse que pendant la période d'émergence du champ « *l'identité de l'entité qui deviendra plus tard une innovation (si tel est le cas) est fondamentalement ambiguë, ambiguïté sur laquelle jouent les innovateurs pour la faire progressivement accepter et sur laquelle s'appuient ceux qui la reçoivent pour l'approprier ou, au*

³¹⁶ Steve MAGUIRE et Cynthia HARDY, « Discourse and Deinstitutionalization: The Decline of DDT », (2009) 52-1 *Academy of Management Journal* 148- 178; D LEVY et M SCULLY, « The Institutional Entrepreneur as Modern Prince: The Strategic Face of Power in Contested Fields », (2007) 28-7 *Organization Studies* 971-991.

³¹⁷ Henri BERGERON, « Eléments pour une sociologie de l'entrepreneur-frontière », (2013) 54-2 *Revue française de sociologie* 263- 302, p. 266.

³¹⁸ *Ibid* p. 272

contraire, tenter de la transformer. »³¹⁹. Ainsi cet entrepreneur frontière ne conduit pas exactement un travail de généraliste qui « *construit un sens commun entre des milieux institutionnels qui ne recouvrent pas aux mêmes savoirs et mêmes représentations* » mais plutôt de courtier cherchant « *des solutions acceptables entre les groupes éloignés qui peuvent trouver un intérêt à coopérer même s'ils ne poursuivent pas les mêmes objectifs et n'ont pas les mêmes intérêts* »³²⁰. Dans son travail de courtier il ajuste ses arguments au gré de ses interlocuteurs et du contexte qui de fait ne passent pas d'accord entre eux mais avec lui.

L'ouverture du pouvoir d'agence à d'autres acteurs que les entrepreneurs contribuant à modifier les objectifs du champ est proposée par d'autres auteurs. Déjà, les travaux de la sociologie de l'innovation considéraient que dans le processus d'innovation l'entrepreneur schumpetérien avait été remplacé par une « *foule d'intervenants diversifiés* »³²¹. Des auteurs comme J. Bartley, C. Hardy et S. Maguire ont souligné l'importance d'élargir le cercle d'acteurs étudiés dans ce processus et de s'intéresser aussi aux acteurs souhaitant défendre leur environnement institutionnel contre la création d'un nouveau champ³²². De plus, le cadre théorique du champ organisationnel et de l'entrepreneur institutionnel avait tendance à dépeindre le processus d'institutionnalisation comme une construction maîtrisée, linéaire, et étudiant surtout *a posteriori* des succès d'institutionnalisation sans développer d'outils pour appréhender les processus n'aboutissant pas à l'institutionnalisation³²³.

En s'appuyant sur ces critiques, T. Lawrence et R. Suddaby ont élaboré le concept de *travail institutionnel* qui désigne « *l'action intentionnelle des acteurs ou des organisations visant à créer, maintenir ou déstabiliser les institutions* »³²⁴. Le concept de travail institutionnel propose dans un premier temps une vision élargie du cercle d'acteurs agissant sur l'institution et une perspective plus modérée de leurs capacités réflexives. Ensuite, il propose une conception de l'émergence d'institution moins linéaire, plus chaotique et étant le fruit d'une interaction entre une variété d'acteurs. Ces travaux s'inscrivent dans la continuité des travaux sur *l'entrepreneur institutionnel* puisqu'ils cherchent à prendre en compte la capacité des acteurs à influencer leur environnement cognitif et institutionnel. Néanmoins, ils proposent une approche plus collective en étendant leur étude sur un plus grand nombre d'acteurs impliqués dans ces processus intentionnels. Ils comprennent l'émergence d'institutions comme un phénomène complexe, résultant d'actions contradictoires, non pas dirigées et maîtrisées par quelques entrepreneurs suivis de façon mimétique par le reste des acteurs³²⁵. S'inspirant des travaux de C. Olivier sur une typologie des réponses stratégiques de résistance aux processus institutionnels, ce concept permet de prendre en compte des

³¹⁹ *Ibid.* p. 293.

³²⁰ Olivier NAY et Andy SMITH, « Les intermédiaires en politique: médiations et jeux d'institutions », dans *Le gouvernement du compromis. Courtiers et généralistes dans l'action politique*, Economica, Paris, 2002, p. 47-86.

³²¹ Madeleine AKRICH, Michel CALLON et Bruno LATOUR, « A quoi tient le succès des innovations? 1: L'art de l'intéressement », *op. cit.* p. 2.

³²² Karim BEN SLIMANE et Bernard LECA, « Le travail institutionnel : origines théoriques, défis et perspectives », (2010) 7-37 *Management & Avenir* 53-69, p.54.

³²³ P.S. TOLBERT et Lynn G. ZUCKER, « The institutionalization of institutional theory », dans Stewart CLEGG, Cynthia HARDY et NORD (dir.), *Handbook of organization studies*, Sage, London, 1996, p. 175-190.

³²⁴ Thomas B. LAWRENCE et Roy SUDDABY, « Institutions and Institutional Work », *op. cit.* p. 215, Nous reprenons ici la traduction de l'anglais : « *the purposive action of individuals and organizations aimed at creating, maintaining and disrupting institutions* » proposée dans, Ben Slimane, Karim, et Bernard Leca. 2010. « Le travail institutionnel : origines théoriques, défis et perspectives ». *Management & Avenir* 7 (37): 53-69,

³²⁵ Karim BEN SLIMANE et Bernard LECA, « Le travail institutionnel : origines théoriques, défis et perspectives », *op. cit.* p. 59.

actions contradictoires à l'œuvre dans le même processus institutionnel³²⁶. Ce concept met en lumière l'activité des acteurs au-delà du résultat de l'institutionnalisation, il caractérise un répertoire de modalités de travail institutionnel pouvant œuvrer dans des sens opposés voir entrer en conflit. Le *travail institutionnel* nécessite une intentionnalité des acteurs, lesquels considèrent l'institution comme une construction sociale pouvant être mise en place ainsi que remise en cause. Il implique une vision de moyen à long terme dont le but est de façonner une norme qui organisera les conditions d'exercice de l'activité dans le champ³²⁷. Le *travail institutionnel* fait référence aussi à l'effort demandé aux acteurs pour s'investir dans cette tâche et s'intéresse aux motivations les conduisant à abandonner l'activité d'ordinaire moins coûteuse de reproduction des routines. Il ouvre un agenda de recherche sur les moteurs du travail institutionnels chez les acteurs.

Si ces auteurs reconnaissent un pouvoir d'*agence* aux acteurs, ils soulignent néanmoins que cette capacité n'est pas issue uniquement d'une décision de libre arbitre, mais est aussi inscrite dans leur appartenance à des structures, des entités de niveau intermédiaire. Ces auteurs reprennent le concept de *pratique* développé par P. Bourdieu, A. Giddens et M. De Certeau, qu'ils comprennent suivant la définition de Schatzki comme une gamme d'activités humaines médiatisées matériellement et incarnées, centralisées autour d'une compréhension pratique partagée³²⁸. Ces *pratiques* s'inscrivent dans des logiques de champs dépassant les contours du *champ organisationnel* ou du *champ d'action stratégique*. Ainsi, le *travail institutionnel* se situe souvent à la rencontre de différents champs, pouvant expliquer pourquoi les acteurs s'engagent dans des stratégies différentes, voire contradictoires, puisqu'ils sont ancrés dans des logiques de champ variables. Quelques acteurs occupent simultanément des positions dans des champs multiples, ils sont à l'intersection et bénéficient des ressources multiples pour amorcer un travail de création institutionnelle³²⁹. L'institution émerge alors par un processus circulaire de co-structuration permanente entre l'acteur et l'institution s'inscrivant explicitement dans les travaux d'A. Giddens. C'est une vision récursive des actions qui sont à la fois contraintes et permises par les structures qui sont produites et reproduites par ces actions³³⁰.

N. Fligstein et D. McAdam s'inscrivent aussi dans la conception de dynamique institutionnelle constructiviste d'A. Giddens³³¹. Leur *théorie générale des champs* propose d'expliquer l'émergence de nouveaux champs par le travail d'*acteurs compétents* (*skilled actors*), en réaction à une crise dans un autre champ. Ces *acteurs compétents* s'apparentent aux *entrepreneurs institutionnels* présentés plus haut, ils détiennent des *compétences sociales*, définies comme des capacités à développer « *des pensées et des actions intersubjectives qui structurent la construction d'un sens, d'intérêts et d'identités au service de fin collectives* »³³². Les *compétences sociales* peuvent être réparties assez également entre les *titulaires* et les *challengeurs*, mais ces derniers seront départagés dans l'entreprise de construction par une répartition des

³²⁶ C. Olivier propose une typologie de 5 actions de résistance stratégiques aux processus d'institutionnalisation de la plus passive à la plus active : Le consentement, le compromis, l'évitement, la défiance et la manipulation. Olivier, C. 1991. « strategic responses to institutional processes ». *Academy of Management Review* 16: p.145

³²⁷ Steve MAGUIRE et Cynthia HARDY, *op. cit.*; Karim BEN SLIMANE et Bernard LECA, *op. cit.* p. 64.

³²⁸ Traduction de l'auteure à partir de l'anglais « *embodied, materially mediated arrays of human activity centrally organized around shared practical understanding* » Schatzki cite dans Thomas B. LAWRENCE et Roy SUDDABY, « Institutions and Institutional Work », *op. cit.* p. 218.

³²⁹ Thomas B. LAWRENCE et Roy SUDDABY, « Institutions and Institutional Work », *op. cit.* p. 248.

³³⁰ Anthony GIDDENS, *Central Problems in Social Theory: Action, Structure and Contradictions in Social Analysis*, University of California Press, 1979.

³³¹ Neil FLIGSTEIN et Doug MCADAM, *A theory of fields*, *op. cit.* p. 26.

³³² Traduction de l'auteure de l'anglais, « *social skill* », « *which we define as the capacity for intersubjective thought and action that shapes the provision of meaning, interests, and identity in the service of collective ends* ». *Ibid* p. 4.

ressources inégales ou des supports politiques en faveur de l'un ou l'autre³³³. Le cercle d'acteurs « actifs » est ainsi plus élargi que dans le cadre du concept *d'entrepreneur institutionnel*.

Ces auteurs s'intéressent aussi particulièrement au contexte de l'émergence des champs. Ils définissent un champ émergent comme « *une arène construite socialement, occupée par deux ou plusieurs groupes dont l'action est orientée les uns envers les autres, mais qui n'ont pas développé un ordre stable capable de rendre routinières les relations du champ* » ; les règles n'existent pas, mais les acteurs, par leurs intérêts dépendants et leurs visions, sont forcés à se prendre en compte les uns les autres dans leur action³³⁴. L'émergence d'un nouveau *champ d'action stratégique* résulte souvent d'un état de crise d'un autre champ et est envisagée par ses promoteurs comme une opportunité pour la réalisation des intérêts d'un individu ou d'un groupe³³⁵. En effet, selon D. North, dans un état d'équilibre institutionnel aucun agent n'a intérêt à consacrer des ressources à des modifications des accords à la base de cet équilibre, le changement est issu d'un choc qui amène une partie d'un accord (ou les deux parties) à percevoir qu'il pourrait gagner d'une modification de l'accord, et donc qu'il pourrait gagner à consacrer des ressources à une renégociation³³⁶. Ces champs orientés vers l'établissement d'une institution sont encastés dans une toile complexe d'autres champs s'influençant les uns les autres.

Afin de mieux comprendre les différences entre ces divers cadres théoriques et de pouvoir justifier nos choix, il est indispensable de revenir sur leur conception des modalités de stabilisation du champ.

2.1.1.2. Etudier le processus de stabilisation des champs et des institutions

Après s'être intéressée à l'impulsion de création du champ, il est nécessaire de clarifier la conception dynamique qu'adopte cette recherche de l'étude de l'évolution du champ. Ce travail s'inscrit dans la tradition des travaux néo-institutionnalistes soulignant la nature incrémentale du changement institutionnel, ainsi que son absence de linéarité et son manque de maîtrise par les acteurs l'instituant.

Les théories néo-institutionnelles du changement s'inscrivent dans l'héritage des théories insistant sur la routinisation des pratiques et la dépendance au sentier au sein des institutions. K. Thelen et S. Steinmo expliquent que, bien qu'on entende souvent par « changement » des transformations radicales des pratiques, au sein des institutions, les changements sont souvent le fruit d'effets cumulatifs de petits changements continus³³⁷. D. North souligne aussi que parmi la diversité de changements institutionnels pouvant être observés, leur particularité commune est de s'effectuer de manière incrémentale, progressive, pas à pas³³⁸. Cette tendance a été d'autant plus soulignée dans le domaine de l'action publique par les travaux de C. Lindblom dans lesquels il décrit les mécanismes menant à l'absence de changement radical dans la programmation politique³³⁹. Cette loi générale peut sembler de prime abord ne pas pouvoir s'appliquer à

³³³ Neil FLIGSTEIN et Doug MCADAM, *A theory of fields*, op. cit. p.54

³³⁴ *Ibid* p.86

³³⁵ Neil FLIGSTEIN et Doug MCADAM, « Toward a General Theory of Strategic Action Fields », op. cit. p. 10.

³³⁶ Douglass C. NORTH, *Institutions, Institutional Change and Economic Performance*, Cambridge University Press, 1990, p. 86-87.

³³⁷ S. STEINMO et K. THELEN, « Historical institutionalism in Comparative Analysis », dans Sven STEINMO, Kathleen Ann THELEN et Frank LONGSTRETH, *Structuring politics: historical institutionalism in comparative analysis*, coll. Cambridge studies in comparative politics, Cambridge [England] ; New York, Cambridge University Press, 1992.

³³⁸ « [...]the single most important point with institutional change (...) is that institutional change is overwhelmingly incremental » Douglass C. NORTH, op. cit. p. 89.

³³⁹ Charles LINDBLOM, « The Science of "Muddling Through" », (1959) 19 *Public Administration Review* 79-88.

l'objet étudié du fait de la rapidité de l'institutionnalisation de la méthode expérimentale en France. Néanmoins, au-delà de l'apparente rupture, ce travail souhaite rester attentif aux points communs qu'elle comporte avec les pratiques précédentes. Il est possible qu'une révolution affichée recouvre finalement une continuité pouvant expliquer en partie la rapidité d'un succès.

Le concept d'*entrepreneur institutionnel* mobilisé par les néo-institutionnalistes semble pertinent pour appréhender l'émergence d'une institution, même si nous avons vu que la toute-puissance de cet acteur pouvait être nuancée. Cependant, ce courant théorique n'apporte pas d'éclairage précis sur le processus de stabilisation du champ nouvellement créé. Comme le soulignent N. Fligstein et D. McAdam à propos de cette conception, il semblerait qu'une fois l'action institutionnelle effectuée, et les nouvelles normes établies, elles soient alors prises pour acquises et qu'une dynamique de routinisation reprenne le dessus³⁴⁰. Selon cette conception, le pouvoir d'agence des acteurs est donc seulement présent pendant une parenthèse, lors de la construction du champ, et contrôlé par un nombre d'acteurs limité. Cette conception n'est pas satisfaisante pour notre objet, de par la malléabilité du champ observé, mais aussi par la multitude d'acteurs influençant son développement.

Au-delà de la dimension incrémentale du changement, les cadres théoriques sollicités dans ce travail amènent le chercheur à considérer le changement et la construction institutionnelle comme un processus non-linéaire et non maîtrisé dans son intégralité par ses instigateurs³⁴¹. Le *travail institutionnel* décrit par T. Lawrence et R. Suddaby est permanent. Il change juste de nature selon la phase de stabilisation de l'institution, allant des activités de création au maintien et enfin à la déstabilisation. Selon cette conception, une institution ne se maintient pas de manière routinière, mais elle est maintenue par le travail permanent d'acteurs.

Ils proposent le répertoire d'actions suivant, en estimant que ces *pratiques* peuvent intervenir de manières concomitantes:

³⁴⁰ Neil FLIGSTEIN et Doug MCADAM, *A theory of fields*, *op. cit.* p. 179.

³⁴¹ Neil FLIGSTEIN et Doug MCADAM, « Toward a General Theory of Strategic Action Fields », *op. cit.* p. 11.

Table 1 : Types de pratiques de travail institutionnel identifiés par T. Lawrence et R. Suddaby

Créer des institutions	Plaider en faveur du projet
	Définir
	Motiver
	Construire des identités
	Changer les associations normatives
	Construire des réseaux d'acteurs sanctionnant le non respect de l'institution
	Coupler les nouvelles pratiques avec d'anciennes
	Elaborer et théoriser les relations de cause à effet
	Fournir aux acteurs les connaissances et les compétences nécessaires pour s'approprier l'institution
Maintenir des institutions	Créer des règles pour soutenir les institutions
	Maintenir l'ordre par le contrôle
	Dissuader
	Valoriser et diaboliser
	Création de mythes autour de l'origine et de l'histoire de l'institution
	Encastrer et rendre routinier
Déstabiliser des institutions	Déconnecter sanctions et pratiques
	Dissocier les pratiques de leurs fondements moraux
	Remettre en cause les présupposés et les croyances

Source : T. Lawrence et R. Suddaby traduit par Ben Slimane et Leca³⁴²

Le *travail institutionnel* étant en partie issu des *pratiques* des acteurs, il est présent dans les activités quotidiennes de ces derniers et, de fait, ne s'arrête pas une fois le processus de création stabilisé. T. Lawrence et R. Suddaby présentent le phénomène d'émergence de l'institution comme résultant de confrontations, de compromis entre les forces de construction et les forces de déstabilisation portées par une diversité d'acteurs et ne correspondant pas à l'intention originelle des entrepreneurs³⁴³. Le travail institutionnel n'est pas linéaire, l'institution n'est pas une variable binaire, présente ou absente, mais elle évolue au gré du travail institutionnel effectué, elle varie en degré d'institutionnalisation et peut même parfois revenir en arrière. D'ailleurs, comme l'a souligné C. Topalov dans son étude du *champ réformateur* constitué au début du siècle entre des scientifiques et des acteurs politiques, certains champs peuvent se constituer autour de la promotion d'une méthode ou d'une réforme de manière temporaire. Un champ réformateur peut ainsi brouiller le comportement des acteurs de champs différents autour de la défense d'un instrument, mais ensuite se dissoudre après son adoption³⁴⁴. Ainsi, bien que cette construction théorique accorde une place de choix à l'initiative des acteurs dans le processus de construction institutionnelle, ce dernier n'est pas complètement maîtrisé. Les acteurs de l'institutionnalisation, et particulièrement les entrepreneurs, tentent d'orienter la construction institutionnelle en faveur de leur cause, à l'aide des ressources variées dont ils disposent sans pour autant exercer un contrôle total sur cette entreprise.

³⁴² Karim BEN SLIMANE et Bernard LECA, *op. cit.*

³⁴³ Thomas B. LAWRENCE et Roy SUDDABY, « Institutions and Institutional Work », *op. cit.* p. 219; Karim BEN SLIMANE et Bernard LECA, *op. cit.* p. 59.

³⁴⁴ Christian TOPALOV, « Le champ réformateur, 1880-1914: un modèle », dans *Laboratoires du nouveau siècle: la nébuleuse réformatrice et ses réseaux en France, 1880-1914*, coll. Civilisations et sociétés, n°98, Paris, Ecole des hautes études en sciences sociales, 1999 p. 463.

Bien que la *théorie des champs* ne se réfère pas au concept de *travail institutionnel*, elle relève les mêmes écueils à propos du concept d'*entrepreneur institutionnel* et propose une conception similaire à celle de T. Lawrence et R. Suddaby. N. Fligstein et D. McAdam conçoivent le champ comme un flux, en mouvement constant, provoqué à la fois par les acteurs au sein du champ et par leurs relations avec les champs voisins avec lesquels ils sont connectés. Le processus de contentieux est présent en permanence entre les *titulaires* et les *challengeurs*, il varie juste en termes de degré selon l'identification d'opportunités de renverser les titulaires par un groupe de *challengeurs* dotés de *compétences sociales*³⁴⁵. Ces auteurs évoquent une danse stratégique itérative, où chacun modifie constamment sa stratégie et ses tactiques en réponse aux mouvements de l'autre³⁴⁶. Ils défendent ainsi une conception du changement s'incarnant d'une part par un dynamisme incrémental permanent ainsi que des ruptures occasionnelles provoquées à l'intérieur du champ, et d'autre part pouvant être issu d'un champ à proximité, d'un événement à un niveau macro (crise économique par exemple) ou d'une invasion par un groupe extérieur. Les *titulaires* du champ ont donc pour souci constant de stabiliser leur position. Ils le font par la diffusion d'un discours « *existentialiste* », c'est à dire d'appartenance au champ orienté en leur faveur, en cherchant un soutien auprès des champs stratégiques d'État, en constituant des *unités de gouvernance interne* dont le but est de stabiliser les accords du champ, renforcer la perspective dominante, garder les intérêts des titulaires, naturaliser la logique et les règles du champ et servir de liaison avec d'autres champs extérieurs. Cette recherche n'étudie pas un champ émergent ayant vocation à se transformer en institution. Elle cherche à comprendre la dynamique d'émergence d'un nouveau champ ayant pour vocation l'établissement d'une nouvelle institution, ici un instrument d'action publique. Ce chapitre s'intéresse au processus chaotique de promotion de la méthode expérimentale par une variété d'acteurs chacun influençant sa forme et le discours légitimateur l'environnant afin de rassembler le maximum de soutien.

2.1.1.3. Les variables d'intérêt pour appréhender l'émergence d'une institution

Après avoir identifié les acteurs de l'institutionnalisation ainsi qu'explicité notre perception du mouvement de la construction institutionnelle, l'objet principal de ce travail consiste à s'intéresser aux pratiques et stratégies développées par ces acteurs pour participer à cette construction. Tous les différents cadres théoriques présentés ci-dessus ont en commun d'accorder une place de choix au travail cognitif exercé par et sur les acteurs du champ, cette étude occupe le centre de cette recherche. Ce travail est d'autant plus central du fait que le *champ d'action stratégique* étudié est composé d'acteurs scientifiques engagés dans des activités quotidiennes de production cognitive. Ce répertoire d'outils cognitifs est aussi complété par d'autres leviers présents dans la littérature néo-institutionnaliste, comme la création de normes, de régulation, et l'allocation de ressources.

La *sociologie de la traduction* de M. Callon aborde l'innovation comme un processus mené par un ensemble d'entrepreneurs scellant des alliances autour, d'abord, d'un travail cognitif³⁴⁷. Ce travail de *problématisation* a pour but de mobiliser les alliés indispensables à la réussite de l'innovation en formulant leur intérêt dans ce processus. Dans un premier temps, les premiers entrepreneurs formulent et mettent en scène les

³⁴⁵ Neil FLIGSTEIN et Doug MCADAM, *A theory of fields*, op. cit. p. 12.

³⁴⁶ *Ibid* p.84

³⁴⁷ Madeleine AKRICH, Michel CALLON et Bruno LATOUR, « A quoi tient le succès des innovations? 1: L'art de l'intéressement », op. cit. p. 5.

problèmes rencontrés par ces acteurs dans l'atteinte de leurs buts. Ces problèmes, différents pour chaque groupe d'acteurs, sont alors synthétisés en une problématique commune. L'innovation est alors présentée comme un moyen de résoudre cette problématique, un *point de passage obligé* permettant de surmonter l'obstacle identifié de manière profitable pour chacun d'eux³⁴⁸.

On retrouve ce même intérêt pour le travail cognitif effectué par des entrepreneurs au cours du processus d'innovation dans les travaux du néo-institutionnalisme sociologique, et notamment dans le concept de *champ organisationnel*. Le rôle des *entrepreneurs institutionnels*, comme dans la sociologie de la traduction, est de participer à la formulation et définition d'un problème commun, tout en proposant une solution pour le résoudre selon leurs propres enjeux et intérêts³⁴⁹. La construction d'un champ institutionnel se manifeste par l'adhésion à un mythe commun rationalisé - le problème identifié et la façon d'y remédier - qui produit un système de croyances délimitant des cadres d'action. Le concept de *champ réformateur*, évoqué plus haut, porte le travail cognitif au cœur du processus de constitution du champ commun entre acteurs politiques et scientifiques, dont le but est « *l'énoncé de problèmes, de diagnostics et de prescriptions, bref, un langage et une méthode* »³⁵⁰. C'est une entreprise de production de sens et sa diffusion comme sens commun. Si l'on se réfère au répertoire des types de travail institutionnel établi par T. Lawrence et R. Suddaby, on retrouve aussi cette prégnance des stratégies de natures cognitives. Si l'on se reporte au tableau présenté ci-dessus, on note que l'intégralité des actions de création mobilise un travail cognitif ainsi que la plupart des actions de déstabilisation ou de maintien. L'étude des processus et structures cognitives à l'œuvre est au cœur de l'agenda de recherche ouvert par ce cadre théorique³⁵¹.

Les travaux de l'école cognitive française de l'analyse des politiques publiques relayent aussi cet intérêt pour ce type d'activités. La notion de *mythe* permet d'appréhender les variables cognitives dans leur diversité, mais néanmoins réunies dans un récit plus ou moins cohérent guidant l'action des acteurs. Comme le soulignent F. Desage et J. Godard, sa grammaire est suffisamment imprécise pour que les acteurs y voient leurs propres attentes, représentations et intérêts. On peut rapprocher cette notion de *mythe* de celle d'objet frontière que nous avons présenté plus haut. Ces auteurs conçoivent les mythes comme des « *fétiches rassembleurs d'initiatives variées qui préservent par leur ambigüité constitutive une pluralité d'options dans lesquelles les bricolages, les routines, mais également les confrontations d'intérêts vont pouvoir prendre place* »³⁵². La notion de mythe peut donc s'articuler avec une conception de la construction institutionnelle comme un processus non linéaire puisqu'il permet d'étudier l'hétérogénéité de points de vue trouvant néanmoins chacun leur intérêt dans un discours général porté, pour notre cas, sur l'action publique. Ainsi, le travail cognitif élaboré par les entrepreneurs institutionnels contribue à cadrer et limiter les possibles marges d'actions des *challengers* adhérents ou opposants à cette construction institutionnelle, suivant le principe de la rationalité limitée développé par J. March et J. Olsen.

³⁴⁸ Michel CALLON, « Eléments pour une sociologie de la traduction, la domestication des coquilles Saint-Jacques et des marins-pêcheurs dans la baie de Saint-Brieuc », (1986) 36 *l'Année sociologique* pp 169-208, p.183.

³⁴⁹ M. C. SUCHMAN, « Managing Legitimacy: Strategic and Institutional Approaches », (1995) 20 *Academy of Management Review* 571-610; R. DELBRIDGE et T. EDWARDS, « Challenging conventions: Roles and processes during non-isomorphic institutional change », (2008) 61-3 *Human Relations* 299-325.

³⁵⁰ Christian TOPALOV, *op. cit.* p. 471.

³⁵¹ Thomas B. LAWRENCE et Roy SUDDABY, « Institutions and Institutional Work », *op. cit.* p. 218.

³⁵² Fabien DESAGE et Jérôme GODARD, « Désenchantement idéologique et réenchantement mythique des politiques locales », (2005) 55-4 *Revue Française de Science Politique* 633-661, p.658.

Au-delà d'être rassembleur, l'avantage du mythe est de pouvoir formuler une causalité, c'est-à-dire une mise en récit de son action *a posteriori* selon des idées, et contribuer à la légitimation de l'action publique. Il revêt aussi l'avantage de constituer une croyance vraisemblable et infalsifiable, le mythe n'a pas besoin d'être validé, il échappe à la raison scientifique comme à l'expérience pratique³⁵³. Il peut exposer des éléments contradictoires, sa force est de « *mettre en récit* » l'action publique sans contrainte de résultats. Comme le souligne P. Krugman, beaucoup de croyances politiques en économie viennent d'anecdotes, de petites histoires se distinguant d'un raisonnement scientifique³⁵⁴. Cette notion de mythe semble a priori difficilement applicable à l'objet de cette étude étant donné que le processus étudié promeut l'adoption d'une démarche scientifique dans la conduite de l'action publique. Il semblerait paradoxal qu'une démarche de promotion de l'*evidence-based policy* puisse être constituée sur une mise en récit échappant elle-même à une logique scientifique de démonstration. Ce travail sera néanmoins attentif à ce phénomène de mise en récit, de problématisation de l'action, en tâchant d'identifier le discours environnant la promotion de cette méthode, ainsi que les moyens développés pour étayer cette promotion argumentée.

L'attention accordée au pouvoir d'agence des acteurs dans ces cadres théoriques nous amène à ne pas penser les variables cognitives uniquement comme des variables indépendantes, mais de s'intéresser aux interactions entre acteurs ayant mené à leur formulation. S'ouvre alors un programme de recherche consistant à partir à la « *chasse aux producteurs de mythe* », ces personnes ou organisations mobilisant des alliés pour assurer le succès de l'innovation³⁵⁵.

Au-delà de cette attention pour la circulation des idées et l'élaboration de mythes fondateurs, d'autres types d'activités sont présentes dans l'entreprise institutionnelle. Ce travail sera attentif notamment :

- à la création ou suppression de normes, assorties de sanctions formelles ou non pour faire respecter les règles du champ ;
- à l'allocation de ressources à la fois financières et symboliques entre les individus des champs ;
- aux relations interpersonnelles, entre les individus au sein du champ, mais aussi à l'extérieur du champ, pouvant expliquer les stratégies au sein de ce dernier.

2.1.1.4. Opérationnalisation de la construction théorique

Au vu de ces différents cadres théoriques, et au regard de la nature de notre objet de recherche, nous retenons le programme de recherche et les concepts suivants pour étudier l'émergence de la méthode expérimentale par assignation aléatoire en France.

L'articulation des concepts de *travail institutionnel* et de *champ d'action stratégique* permet d'étudier l'émergence de la méthode expérimentale en procédant dans un premier temps à l'identification des champs dont la rencontre est organisée par le *travail institutionnel* visant à institutionnaliser cette méthode comme instrument légitime d'action publique. Il convient alors, de trouver les motivations ayant mené certains acteurs de ces champs distincts à créer un nouveau *champ d'action stratégique* commun autour de la promotion de cette méthode, de comprendre les crises intervenant dans les

³⁵³ *Ibid* p. 656

³⁵⁴ *Ibid* p. 657

³⁵⁵ *Ibid*. p. 659.

champs auxquels ils appartiennent et en quoi la constitution de ce *champ d'action stratégique* constitue une réponse à ces crises.

Au vu des premiers constats émanant du terrain et de ces éclairages théoriques, nous avons formulé la première hypothèse suivante permettant d'expliquer pourquoi un travail institutionnel a été amorcé autour de la promotion de cette méthode.

Hypothèse 2.1 : L'évaluation expérimentale par assignation aléatoire a constitué une ressource pour faire face à des crises autant dans le champ de l'action publique que pour celui de la science économique appliquée en accordant de nouveaux avantages de positionnement à ses promoteurs

Afin de vérifier cette hypothèse, nous identifierons le contenu du discours du *mythe fondateur* entourant la promotion de cette méthode et nous analyserons à quelles crises, et de quels champs, il fait écho. Nous comprenons le terme de crise selon la définition de M. Dobry comme « un état du système social dont la caractéristique essentielle est la fluidité conjoncturelle des rapports sociaux »³⁵⁶. Cette période se caractérise par une sortie d'une conjoncture routinière, un moment où les rôles ainsi que les secteurs se brouillent, où les acteurs se trouvent dans une incertitude structurelle puisque qu'ils n'ont plus les moyens routiniers d'anticipation et d'appréciation des situations. C'est un moment où il est nécessaire de redéfinir le monde social, puisque la conjoncture de crise réduit les définitions opérationnelles en période routinière. L'objet de cette recherche n'est pas la description de la réalité de ces crises mais plutôt l'identification de l'usage de la grammaire autour de la conjoncture de crise et la présentation de cette méthode comme une solution. Nous nous interrogerons sur la présence d'ambiguïtés dans ce discours ainsi que sur les identités de ces promoteurs, et la malléabilité de ce discours lors du processus de promotion. Les identités des entrepreneurs de cette méthode nous orienteront vers l'étude des champs susceptibles d'être liés avec ce nouveau champ émergent. Nous nous interrogerons, sur les relations qu'entretiennent ces champs entre eux en mobilisant les outils de la *théorie générale des champs*. Nous nous demanderons, s'ils sont proches ou distants, si leur relation est horizontale ou verticale, ainsi que sur la nature étatique de ces derniers.

Nous nous demanderons ensuite, pourquoi cette méthode-ci en particulier a été choisie comme instrument de *l'evidence-based policy* en France. Nous faisons l'hypothèse que la seule logique d'importation étrangère de ce modèle ne peut constituer la principale explication. Les échecs précédents des tentatives de diffusion de pratiques d'évaluation étrangères au sein de l'administration centrale témoignent de la difficulté à importer ces instruments dans le contexte français. Nous formulons ici l'hypothèse que :

Hypothèse 2.2 : Cette méthode doit son succès en France, à sa mobilisation des acteurs traditionnels de l'expertise de l'administration centrale et s'inscrit en partie dans une logique prospective et non rétrospective correspondant à la logique française de mobilisation des savoirs scientifiques

³⁵⁶ Michel DOBRY, *Sociologie des crises politiques la dynamique des mobilisations multisectorielles*, 3e édition, Paris, Presses de Science Po, 2009.

Nous estimons ici que dans un premier temps, la méthode expérimentale n'a pas été un instrument de reconfiguration des interactions entre les acteurs de l'expertise publique et ceux de l'action publique, au contraire, elle permettait justement dans un premier temps d'avoir recours à une méthode d'évaluation conduite par des interlocuteurs privilégiés de l'action publique en France, praticiens participant déjà en interne à l'expertise publique et inscrits dans une logique d'aide à la décision et non de jugement de l'efficacité. En effet, comme nous l'avons vu précédemment cette méthode porte principalement sur de nouveaux types d'intervention et ne menace pas la légitimité des interventions existantes. Approfondir ces pistes, nous permettrait de comprendre pourquoi, contrairement aux autres initiatives précédentes de promotion de l'évaluation au sein de l'administration centrale, cette méthode fut si rapidement diffusée.

Nous étudierons ainsi, les modalités traditionnelles d'interaction entre le champ de l'action publique et de l'expertise d'Etat en France et l'habileté de la méthode expérimentale par assignation aléatoire à s'inscrire dans cet héritage contrairement aux autres méthodes d'évaluation.

Une fois que nous aurons analysé, d'une part, les structures sous-jacentes à l'interaction des sciences sociales et de l'action publique au sein de l'Administration centrale française (*Hyp. 2.2*), ainsi que les motivations pouvant expliquer l'organisation d'un travail institutionnel (*Hyp. 2.1*) autour de la constitution d'un *champ d'action stratégique* promouvant cette méthode, nous aborderons alors, le travail d'agence des promoteurs de cette méthode et du courant de *evidence-based policy* en France. Notre étude préalable de la diffusion américaine de cette méthode, ainsi que le programme de recherche sur les instruments d'action publique appellent à ouvrir une piste de recherche sur l'identification des promoteurs de cette méthode, leur appartenance, leur organisation et leurs modalités d'action. Le développement précédent a montré l'importance déterminante de certains acteurs ou organisations dans la définition des instruments de l'EBP en soulignant aussi le contact qu'elle nécessite dans cette entreprise entre acteurs de l'action publique et de la production scientifique. La disparition actuelle de la promotion explicite de cette méthode par l'administration centrale nous amène à envisager que cette alliance n'a pu s'opérer que dans une configuration particulière aujourd'hui modifiée. Nous formulons donc l'hypothèse que :

Hypothèse 2.3 : La promotion de cette méthode a bénéficié d'un travail institutionnel intense conduit par un champ d'action stratégique temporaire alliant acteurs scientifiques et acteurs de l'action publique

La vérification de cette hypothèse implique d'abord un travail d'identification des acteurs de ce champ, ainsi que la délimitation de ce dernier.

En croisant les typologies proposées par T. Lawrence, R. Suddaby, N. Fligstein et D. McAdam, nous considérons la possibilité d'identifier 4 types d'acteurs engagés dans cette entreprise. Ils s'organisent selon une matrice à deux variables, la première porte sur la nature du travail institutionnel, à savoir de construction ou de déstabilisation et la seconde sur le niveau des compétences sociales relatives au champ détenues par les acteurs.

- D'abord, nous nous intéresserons aux *titulaires*, aux entrepreneurs principaux qui ont impulsé de manière significative ce mouvement. Nous interrogerons leur profil, les dotant de *compétences sociales* particulièrement pertinentes pour ce champ et tenterons de caractériser leur rôle. Nous interrogerons la présence d'entrepreneur-frontière à identité

modulable et nous étudierons leur entreprise de structuration du champ autour notamment, de la création *d'une unité interne de gouvernance* du champ.

- Nous nous intéressons aussi aux autres acteurs, que nous caractériserons de *suiveurs*, ce sont les autres individus, ayant contribué par leur pratique, par un travail moins intense, à consolider ce *champ d'action stratégique* dans le sens des *titulaires*. Ces acteurs peuvent articuler un discours un peu différent, voire être de potentiels *challengeurs* si une occasion se présente. Ils sont sensiblement moins dotés en compétences sociales vis-à-vis de ce champ particulier.
- Nous chercherons ensuite à identifier les *challengeurs compétents*, qui pratiquent un travail de déstabilisation du champ, et sont dotés de *compétences sociales* significatives. Ceux-ci peuvent s'inscrire par nécessité dans le champ, de par leur situation de dépendance vis-à-vis de ce dernier, ou être encore inscrits dans d'autres champs interreliés avec ce dernier. Nous tenterons de comprendre pourquoi ces individus se placent en *challengeurs* et s'engagent dans des actions de déstabilisation en identifiant la nature de leurs pratiques.
- Enfin, nous porterons une attention moindre aux *challengeurs* non compétents socialement pour ce champ, ne dérangeant pas intentionnellement *le champ d'action stratégique*, mais perturbant l'action des titulaires par leurs *pratiques* issues d'autres champs et encore non adaptées à ce nouveau champ.

Ces typologies doivent être considérées non pas comme des identités, mais des rôles, que les acteurs peuvent endosser tour à tour au gré des modifications de l'allocation de ressources au sein du champ ou des modifications d'environnement issues des mouvements des autres champs interreliés. La matrice ci-dessous reprend cette typologie

Table 2 : Typologie d'acteurs selon leurs compétences sociales et la nature de leur travail institutionnel

	Travail de création et de maintien	Travail de déstabilisation
Acteurs compétents (<i>Skilled</i>) relativement au champ	Titulaire (<i>crée et maintient le champ par des actions significatives</i>)	Challengeur compétent (<i>dérange le champ par des actions significatives</i>)
Acteurs moins compétents relativement au champ	Suiveur (<i>reproduit le champ par ses pratiques</i>)	Challengeur non compétent (<i>dérange le champ par ses pratiques</i>)

Une des tâches de cette recherche sera ainsi d'établir une cartographie des acteurs ayant participé à l'élaboration du nouveau *champ d'action stratégique*, contribuant à la promotion de cette méthode, en qualifiant leurs différents rôles ainsi que leurs appartenances à des champs multiples. Ce travail permettra aussi de comprendre les raisons de leur investissement dans ce projet vis-à-vis de leurs champs d'appartenance.

Nous nous intéresserons à l'évolution du champ en termes de délimitation de ses frontières, mais aussi en termes de contenu cognitif, de définition de son identité, et de portée *existentielle* pour reprendre les termes de N. Fligstein et D. McAdam. Nous questionnerons aussi la possible stabilisation de ce dernier.

Ce travail de recherche tente de démontrer la pertinence de l'application de ces cadres théoriques à l'étude de l'émergence de l'utilisation de la méthode expérimentale en France. Ces cadres influencent les variables auxquelles la collecte de donnée a été attentive en suggérant un programme de recherche orienté vers l'étude des mécanismes d'alliance, des positions stratégiques des acteurs, tout en accordant une place majeure à la compréhension de leurs propres cadres cognitifs ainsi que ceux qu'ils véhiculent. Ce travail permet une confrontation empirique de ces cadres, pour rendre compte de l'émergence d'une institution prenant la forme ici d'un instrument d'action publique. Les auteurs des différents cadres théoriques présentés ne précisent pas explicitement les méthodologies de collecte empirique leur correspondant. N. Fligstein et D. McAdam évoquent la possibilité d'employer l'analyse de réseau pour cartographier les champs, mais ne suggèrent aucune méthodologie ni épistémologie particulière. Le développement suivant propose de préciser les choix effectués pour mener à bien ce travail empirique en correspondance avec cette élaboration théorique.

2.1.2. Démarche empirique du chapitre 2

Afin de renseigner ces hypothèses de recherche et les différentes dynamiques étudiées, nous mobilisons des outils de collecte et d'analyse de données variés. Nous proposons une analyse croisée de données, issues d'une analyse documentaire mobilisant autant la littérature académique que la littérature grise et approfondie par des entretiens auprès des principales parties prenantes du phénomène étudié. Enfin, nous avons aussi eu recours à une analyse de réseau, nous offrant une représentation graphique des personnes ayant contribué au développement de cette méthode et nous permettant d'établir leur position dans ce nouveau champ.

2.1.2.1. Une étude documentaire de production scientifique, administrative et politique éclairée par le récit des parties prenantes

La collecte de données empiriques a mobilisé des outils de recueil de données primaires et secondaires. L'analyse documentaire et les entretiens semi-directifs ont été conduits en parallèle entre 2010 et 2013.

La collecte de données secondaires a ciblé particulièrement l'ensemble des documents de littérature grise disponibles de 2004, année de la première intervention d'Esther Duflo présentant cette méthode sur le territoire français, à 2011, soit un an après le départ de Martin Hirsch du gouvernement. La littérature grise se compose des textes de lois, de discours, de rapports, de guides, de communiqués de presse, de rubrique de sites internet, de tout type de documents apportant un éclairage sur l'objet de recherche. Cette littérature a aussi été complétée par une étude de l'activité scientifique liée au développement de cette méthode, comprenant des articles de revues scientifiques à comité de lecture, articles de vulgarisation, actes de colloques, curricula scientifiques des acteurs. Le corpus pertinent a été identifié d'une part par des recherches internet par mots-clés, et d'autre part lors des entretiens auprès des personnes interrogées. Pour cette analyse documentaire, nous sommes remontés jusqu'au début des années 1980, afin d'identifier les origines du développement de cette méthode.

La collecte de données secondaires a été complétée par une collecte de données primaires, par l'observation participante et la conduite d'entretiens semi-directifs. La dimension d'observation participante a déjà été évoquée en introduction. Intervenant en amont, et pendant le travail de thèse, notre immersion dans le milieu de l'évaluation a permis d'organiser, en dehors de l'entretien, un "rapport immédiat" avec l'objet de

recherche, et notamment de s'imprégner de l'univers étudié³⁵⁷. Néanmoins, ce premier accès au terrain nécessitait d'être complété par des entretiens semi-directifs mobilisés à titre informatifs et compréhensifs.

Au total, 45 entretiens d'une heure trente en moyenne ont été conduits, soit en face à face, soit par téléphone - du fait de l'éloignement géographique -, auprès de chercheurs, principalement en science économique, et d'acteurs de l'action publique (voir liste en annexe 2). Comme l'ont souligné M. Bachir et V. Bussat, ces entretiens ne peuvent être classés dans une dichotomie entre entretiens directifs et non directifs³⁵⁸. Ils étaient semi-directifs, au sens où une partie de l'entretien se rapprochait d'une démarche directive et visait à réunir des éléments informatifs, mais, en pratique, une partie importante de chaque entretien comprenait une dimension non directive et compréhensive laissant place aux représentations des acteurs.

Les entretiens étaient employés à titre informatif, puisqu'ils constituaient le seul outil permettant de compléter les lacunes des sources écrites à disposition pour reconstruire l'histoire de cet objet d'étude, ainsi que les réseaux de relations et de connaissances interpersonnels entre les acteurs étudiés.

Ils étaient aussi employés à titre compréhensif afin de reconstruire les modèles culturels, les représentations, les systèmes de significations, de valeurs, à travers lesquels les acteurs étudiés perçoivent ou percevaient leur rôle et leur action³⁵⁹. Cet outil était indispensable pour renseigner la dimension cognitive de cette étude et, notamment, les variations de modèles culturels entre les pratiques et représentations des acteurs de l'action publique d'une part, et celles des scientifiques d'autre part, ainsi que leur adhésion et leur participation à la diffusion du discours unifié et de représentations communes autour de la méthode expérimentale³⁶⁰. Cette perspective webérienne, visant à comprendre l'activité sociale par son interprétation, présuppose que l'acteur détient une connaissance du social et que l'entretien est un moment de "conscience réflexive"³⁶¹. Les enquêtés sont ici considérés, pour reprendre les termes de J. C. Kaufmann, comme des « informateurs » qui nous livrent, dans une économie de temps et de moyens, des éléments que l'on aurait pu difficilement recueillir par une étude ethnographique.

Ces entretiens ont été réalisés sur une période d'un an et demi, d'avril 2012 à juillet 2013, auprès d'acteurs clefs identifiés lors de l'analyse documentaire dont la liste a été complétée par une démarche en « boule de neige », c'est-à-dire qu'il était demandé systématiquement à chaque acteur, dans la mesure du possible, de conseiller des personnes qu'ils estimaient avoir été des acteurs clés dans le processus, ou détenant une expertise sur ce sujet³⁶². Cet apport a été déterminant pour ce travail de terrain qui mobilise le savoir et le rôle d'acteurs des « coulisses » de l'État, et de hauts

³⁵⁷ Myriam BACHIR et Virginie BUSSAT, « L'entretien en actes », dans Myriam BACHIR (dir.), *Les méthodes au concret: démarches, formes de l'expérience et terrains d'investigation en science politique*, Presses Universitaires de France, 2000, p. 31-58 p. 33.

³⁵⁸ *Ibid* p. 31

³⁵⁹ Jean-Marie DONEGANI, Sophie DUCHESNE et Florence HAEGEL, « Sur l'interprétation des entretiens de recherche », dans *Aux frontières des attitudes entre le politique et le religieux: textes en hommage à Guy Michelat*, coll. Collection Logiques politiques, Paris, L'Harmattan, 2002, p. 273-295 p. 275.

³⁶⁰ Philippe BONGRAND et Pascale LABORIER, « L'entretien dans l'analyse des politiques publiques: un impensé méthodologique? », (2005) 55-1 *Revue Française de Science Politique* 73-111, 2005.

³⁶¹ Daniel Bertaux cité par Gilles PINSON et Valérie SALA PALA, « L'entretien en sociologie de l'action publique », (2007) 57-5 *Revue Française de Science Politique* 555-597, p.588.

³⁶² Cette méthode dite « snowball » est utilisée notamment par la fondation Cochrane dans la conduite de ses revues systématiques pour s'assurer d'avoir accès à un maximum de documents pertinents, publiés ou non, voir le guide Cochrane en ligne pour davantage d'information <http://handbook.cochrane.org/>

fonctionnaires qui constituent un groupe dont le rôle n'est pas visible au départ³⁶³. La campagne d'entretiens a été arrêtée quand aucune nouvelle personne n'était mentionnée lors des entretiens et que les données collectées arrivaient à "saturation", c'est-à-dire que très peu de nouveaux éléments enrichissant l'analyse étaient recueillis lors de l'entretien.

L'utilisation de cette méthode comporte plusieurs biais potentiels. D'abord, le recours à l'entretien semi-directif dans une collecte d'informations rétrospectives pose la question de la fiabilité de la mémoire des acteurs, puisque les images ne sont pas « stockées » dans la mémoire, mais bien « *retravaillées, reconstruites et modelées en fonction de la perception de la situation du moment* »³⁶⁴. De plus, au-delà de cette distorsion de la mémoire, il est envisageable que certains interlocuteurs, notamment les personnes de pouvoir, aient été dans une stratégie de dissimulation de certaines pratiques. H. Becker nous conseille de douter « *de tout ce qu'une personne de pouvoir peut vous dire* », d'autant plus que celle-ci sans être dans une démarche de dissimulation a parfois des difficultés à verbaliser ses propres pratiques³⁶⁵. Ce risque de l'entretien touche particulièrement cette recherche, puisqu'elle a pu être aussi déclenchée par un contexte d'énonciation particulier, notamment lié à la relation d'intersubjectivité à l'œuvre dans l'entretien. Comme précisé précédemment, l'immersion dans le milieu de l'évaluation nous a amenée à être parfois connue au préalable par nos interlocuteurs, et parfois perçue comme "suspecte". Certains interlocuteurs, notamment les acteurs scientifiques défendant l'utilisation de cette méthode pouvaient parfois afficher une attitude défensive, nous demandant quelle était notre position personnelle vis-à-vis de cette méthode. De manière générale, cette recherche, pour reprendre les mots d'A. Desrosières, a eu le souci de ne jamais considérer les données comme "données" mais plutôt comme des "prises", teintées d'un cadrage et pour lesquelles l'intersubjectivité à l'œuvre entre l'interviewé et l'intervieweur constituait acteur à part entière³⁶⁶.

Ces biais potentiels intervenant dans la collecte de données ont conduit à une analyse ayant comme préoccupation à la fois, de croiser les sources d'éléments empiriques informatifs pour contrôler le récit factuel, mais aussi de travailler sur les interprétations subjectives parfois déformatrices des faits de ces acteurs pour accéder à leurs systèmes de représentations.

2.1.2.2. Les analyses des données : triangulation, analyse compréhensive, et analyse de réseau

Les collectes de données, à titre informatif, pour reconstruire le développement historique de notre objet d'étude ont été traitées par triangulation. En d'autres termes, un crédit a été accordé, en priorité, aux éléments empiriques corroborés par des sources multiples. La triangulation a été double. Elle a consisté à croiser d'une part plusieurs sources utilisant le même outil de collecte de données, comme la parole de plusieurs interlocuteurs sur le même sujet par exemple, et d'autre part, à croiser l'information issue de plusieurs outils de collecte de données, comme l'analyse de document scientifique, l'entretien et l'observation d'un séminaire par exemple.

Les données collectées pour informer sur les représentations des acteurs étaient présentes autant dans les entretiens que dans les productions écrites des acteurs,

³⁶³ Philippe BONGRAND et Pascale LABORIER, *op. cit.* p. 77.

³⁶⁴ J. Peneff cité par Gilles PINSON et Valérie SALA PALA, *op. cit.* p. 558.

³⁶⁵ Howard BECKER, *Les ficelles du métier*, coll. Guide repère, La Découverte, 2002, p. 153-154.

³⁶⁶ Alain DESROSIERES, *L'argument statistique II, Gouverner par les nombres*, coll. Sciences sociales, Paris, Presses de l'école des mines, 2008, p. 14.

qu'elles soient scientifiques ou administratives. Cette collecte de données à visée compréhensive, donnant accès au langage des acteurs, à leurs représentations et interprétations, a fait l'objet d'une analyse à part. Nous avons ici recours à la méthode de Guy Michelat, qui s'inspire de l'analyse des mythes de C. Lévi-Strauss, procédant par une imprégnation du discours et formulant des hypothèses interprétatives de structures, provisoires, malléables au cours de l'analyse³⁶⁷. Dans une démarche interprétative, l'apport du chercheur est principalement la mise en écho des discours entre eux, comme le précisait E. Hugues cité par H. Becker, « *il n'y a rien que je sache qu'au moins un des membres de ce groupe ne sache également, mais, comme je sais ce qu'ils savent tous, j'en sais plus que n'importe lequel d'entre eux* »³⁶⁸.

En pratique, la méthode de Guy Michelat consiste à :

- pour chaque entretien, suite à la retranscription de l'entretien, retenir toutes les propositions qui expriment des significations autonomes ;
- relier chaque proposition aux autres par un trait avec laquelle il a un lien de signification ;
- ajuster ces liens et garder les plus heuristiques pour cette recherche ;
- mettre en parallèle les analyses des entretiens entre eux et élaborer de nouvelles versions avec des propositions qui se répondent et, à partir de là, dégager des modèles d'interprétation.

Ce travail autour des entretiens a été complété par l'analyse de documents constituant un apport sur les interprétations des acteurs. Nous avons mis en écho autant des données factuelles - comme le parcours professionnel et académique de l'interlocuteur - que des données interprétatives, ainsi que le vocabulaire choisi.

Cette démarche inductive a permis à cette recherche de garder une analyse ouverte qui ne s'inscrivait pas uniquement dans la vérification de l'applicabilité d'un cadre théorique, mais qui interrogeait régulièrement la pertinence de ce cadre et pouvait amener aussi à en changer si cela s'avérait plus pertinent au vu de l'analyse intermédiaire des données.

Enfin nous avons eu recours à une analyse de réseau, afin de cartographier les liens entre les acteurs étudiés dans ce chapitre et leur évolution dans de temps, tout en caractérisant la place de chaque acteur dans ce champ émergent. Nous présenterons cette dernière dans le chapitre 2. Cet échafaudage théorique et méthodologique établi pour le chapitre 2, nous pouvons procéder maintenant à la présentation du même travail pour le chapitre 3.

2.2. Appréhender la méthode randomisée dans sa diversité, une stratégie de recherche néo-institutionnaliste centrée sur l'apprentissage

Dans notre troisième chapitre nous nous intéressons à l'étude des incarnations de la méthode expérimentale par assignation aléatoire et ses capacités à agir comme instrument de *evidence-based policy*. Nous présentons dans un premier temps

³⁶⁷ Cette méthode n'a pas été décrite par G. Michelat mais elle a été développée par ses collaborateurs dans Jean-Marie DONEGANI, Sophie DUCHESNE et Florence HAEGEL, *op. cit.* p. 277.

³⁶⁸ Howard BECKER, *Les ficelles du métier*, *op. cit.* p. 166.

l'élaboration théorique de ce chapitre. Celle-ci reprend les concepts de *travail institutionnel* et de *Champ d'Action Stratégique* en les appliquant à une analyse microsociologique au niveau des relations interpersonnelles. Ces concepts sont articulés autour d'une analyse institutionnelle de l'interaction organisée par la méthode aléatoire, que nous qualifierons ici de *relation clinique*³⁶⁹. Cette *relation clinique* dont le but est d'organiser un apport de connaissances commun aux scientifiques et aux acteurs de l'action publique sera étudiée à travers ses différentes incarnations. Nous nous interrogerons, notamment sur la nature de l'apprentissage proposé par cette méthode d'évaluation, en la situant dans les classifications issues de la littérature de l'apprentissage politique.

Cette recherche propose d'appréhender ce phénomène par une étude des *sites institutionnels* de l'incarnation de cette institution, en considérant qu'une institution n'est pas une structure composée de règles figées dans le temps et l'espace, mais s'incarne de manière différenciée selon les dispositions des parties prenantes à entrer en interaction³⁷⁰.

Une approche par la méthodologie d'étude de cas s'avère la plus fructueuse pour rendre compte de cet objet, au regard de sa contemporanéité, du faible nombre d'incarnations de ce phénomène, de la dimension exploratrice de cette recherche et des difficultés à séparer les éléments contextuels de ceux propres à l'*instrument* étudié³⁷¹. Ces études de cas sont élaborées par une analyse croisée d'une triangulation d'outils de collectes de données (littérature grise, académique, entretiens), elles ont fait l'objet d'un traitement croisé des cas entre eux, par induction analytique, tout en restant inscrit dans l'épistémologie du *réalisme critique*³⁷².

2.2.1. Étudier la méthode randomisée comme un site institutionnel d'une relation clinique organisant un type particulier d'apprentissage

Ce premier développement propose d'aborder cette nouvelle dimension de notre objet de recherche en lui appliquant les cadres théoriques mobilisés précédemment. Notre approche s'inscrit dans le programme de recherche ouvert par les travaux portant sur les *instruments d'action publique*. Cette approche sociologique, qui étudie l'impact des outils mobilisés par les acteurs de l'action publique sur l'organisation de leurs interactions, a déjà fait l'objet de nombreux travaux dans le champ de la recherche évaluative et de la sociologie de l'action publique. Nous précisons ici, quelles constructions théoriques sont les plus pertinentes pour notre objet et notre

³⁶⁹ Randall B. RIPLEY, *Policy analysis in political science*, coll. The Nelson-Hall series in political science, Chicago, Nelson-Hall, 1985.

³⁷⁰ Cet auteur définit les sites institutionnels comme « *des espaces sociaux identifiés par trois critères : un ensemble d'acteurs interdépendants (configuration) ; l'existence d'interactions menées en continu (dynamique de jeux) ; et enfin la présence de règles institutionnalisées exprimées dans un ensemble de manières d'agir et de penser (niveau microsociologique) et de logiques sociales (niveau macrosociologiques) pesant sur l'ensemble des comportements et des usages des acteurs en situation de co-présence* ». Nous présenterons plus en détail ce qu'implique l'utilisation de cette notion dans le développement de notre cadre théorique à suivre. Olivier NAY, *La région, une institution: la représentation, le pouvoir et la règle dans l'espace régional*, coll. Collection Logiques politiques, Paris, Harmattan, 1997, p. 20.

³⁷¹ Robert K YIN, *Case study research : design and methods*, Thousand Oaks, Calif., Sage Publications, 2003.

³⁷² Eric W. K. TSANG et Kai-Man KWAN, « Replication and Theory Development in Organizational Science: A Critical Realist Perspective », (1999) 24-4 *The Academy of Management Review* 759-780.

problématique. Chaque expérimentation est ici abordée comme un *site institutionnel*, où s'incarne de manière différenciée, selon les dispositions des parties prenantes à entrer dans les jeux, une institution organisant une *relation clinique* entre deux champs distants dans leurs pratiques d'apprentissage.

2.2.1.1. La méthode randomisée comme un type particulier d'institution d'apprentissage

Ce travail de thèse propose une analyse sociopolitique du transfert d'une méthode scientifique à l'évaluation des politiques publiques. Il appelle à une prise en compte, au-delà des débats épistémologiques, des dimensions d'interactions sociales organisées par cette méthode. Cette élaboration théorique propose de revenir sur l'inscription de cette démarche dans la littérature de l'étude de l'évaluation ainsi que de présenter les liens que cette perspective entretient avec l'approche par les *instruments d'action publique* issue de la sociologie de l'action publique.

La dimension sociale et politique des méthodes d'évaluation

Depuis le milieu des années soixante, les travaux américains de recherche en évaluation s'interrogent sur les méthodes d'évaluations et leur inscription dans leur environnement politique. Ce courant de recherche en évaluation a contribué à définir les pratiques d'évaluations et à les distinguer d'une simple application de méthodes des sciences sociales à l'étude des politiques publiques³⁷³. Ces études autour des pratiques d'évaluation sont regroupées sous les termes de « théories d'évaluations » (*Evaluation Theories*). Selon l'un des anciens présidents de la société américaine d'évaluation, ce sont ces théories qui fondent et distinguent l'identité des évaluateurs³⁷⁴. Deux axes de recherche sont ouverts par cette littérature, d'une part celui de l'étude de l'environnement et des enjeux politiques dans lesquels se conduit l'évaluation, et d'autre part l'étude des méthodologies d'évaluation en pratique et de leurs facultés à favoriser la prise en compte des résultats qu'elles produisent par les décideurs publics.

Ces études évaluatives ont émergé d'un constat de déficit d'utilisation des travaux d'évaluation ou travaux scientifiques par les décideurs publics³⁷⁵. Tout un courant de la recherche évaluative s'est donc intéressé aux vecteurs de l'apprentissage politique pour tenter de proposer un type d'évaluation les prenant en compte et favorisant ainsi l'utilisation de ces résultats³⁷⁶. D'ailleurs, une partie de ces réflexions étaient issues du constat de l'échec de la méthode par assignation aléatoire, dans les années soixante-dix, à favoriser la prise en compte de ses résultats par les décideurs publics³⁷⁷. Ce courant s'appuie sur les hypothèses du modèle de la rationalité limitée de la décision publique élaboré par H. Simon. Il estime que les objectifs de politiques publiques sont rarement clairs, puisqu'ils représentent un consensus entre des intérêts politiques divergents, et que les décideurs publics cherchent rarement à atteindre une vision synoptique du problème, un optimum, mais plutôt à trouver une solution satisfaisante dans le cadre

³⁷³ William R SHADISH, Thomas D COOK et Laura C LEVITON, *Foundations of program evaluation: theories of practice*, Newbury Park, Calif., Sage Publications, 1991, p. 31.

³⁷⁴ William R. SHADISH, « Evaluation Theory is Who We Are », (1998) 19-1 *American Journal of Evaluation* 1-19.

³⁷⁵ Charles LINDBLOM, « The Science of "Muddling Through" », *op. cit.*; Aaron B WILDAVSKY, *Speaking truth to power : the art and craft of policy analysis*, New Brunswick, USA, Transaction Books, 1987.

³⁷⁶ Aux États-Unis, contrairement à la France, on dénombre un nombre important de chercheurs parmi les évaluateurs, ce qui explique l'existence d'un champ de recherche scientifique en évaluation et sur l'évaluation, ainsi que l'existence de plusieurs revues scientifiques consacrées exclusivement à ce champ.

³⁷⁷ Carol H WEISS, « Evaluation Research for Decision: Is Anybody There? Does Anybody Care? », dans Marvin C ALKIN, *Debates on evaluation*, Newbury Park, Calif., Sage Publications, 1990.

d'informations limitées³⁷⁸. De fait, ce courant a formulé l'hypothèse que l'évaluation serait mieux à même d'informer l'action publique si elle était guidée par une meilleure compréhension du champ politique, de la logique guidant les décideurs publics et de la conduite d'une politique publique en général³⁷⁹. L'évaluation ne se restreint pas à l'analyse de la politique publique (*policy*), mais doit alors prendre en compte la politique dans sa pluralité et notamment étudier l'articulation du triptyque : *policy*, en d'autres termes l'action de l'État, *politics*, c'est-à-dire les enjeux de la compétition du pouvoir et *polity* le mode d'organisation de la communauté politique. La méthode choisie lors d'une évaluation devait aussi, dans la mesure du possible, prendre en compte la pluralité des parties prenantes et les intégrer au processus évaluatif afin de susciter leur intérêt pour cette démarche et de favoriser un apprentissage chemin faisant et non uniquement à partir du rendu du rapport final d'évaluation³⁸⁰.

Le terme de *théorie d'évaluation* est susceptible d'être source d'ambiguïté et il nous faut le préciser au préalable³⁸¹. Les auteurs de l'ouvrage *Foundations of Program Evaluation* définissent ainsi ce travail théorique : « *La théorie d'évaluation idéale (jamais atteignable) décrirait et justifierait pourquoi certaines pratiques d'évaluations engendrent un type particulier de résultats au regard des situations auxquelles se confrontent des évaluateurs. Elle (a) clarifierait les activités, les procédures, et buts de l'évaluation ; (b) expliquerait les relations entre les activités évaluatives et les procédures et buts qu'elles facilitent ; et (c) elle testerait empiriquement les propositions d'identifier et d'interpeller les activités évaluatrices qui entrent en conflit avec la recherche ou d'autres connaissances établies sur l'évaluation* »³⁸². Ces mêmes auteurs identifient trois stades d'élaboration de théorie des années 1960 aux années 1990 :

- Les théories de type 1, s'intéressent aux meilleures méthodes pour découvrir la « vérité », dont notamment la théorie de société expérimentale de Campbell³⁸³ ;
- Les théories de type 2, portent sur les usages de l'évaluation et son utilité sociale, et appréhendent l'environnement et l'usage politique de l'évaluation ;
- Les théories de type 3, intègrent les stratégies de recherche évaluative et leur utilité pour la prise en compte des résultats d'évaluation, c'est-à-dire qui font le lien entre les méthodologies d'évaluation et leur prise en compte dans l'environnement politique.

Selon W. Shadish T. Cook et L. Leviton une théorie d'évaluation doit s'intéresser à 5 problématiques :

1. La construction de la connaissance : quelle méthode utiliser pour produire une connaissance crédible ?

³⁷⁸ Nicoletta STAME, « Avantages et inconvénients des différentes méthodes d'évaluation: comment choisir? », dans Sylvie TROSA et COMITE POUR L'HISTOIRE ECONOMIQUE ET FINANCIERE DE LA FRANCE, *Évaluer les politiques publiques pour améliorer l'action publique une perspective internationale*, Paris, Comité pour l'histoire économique et financière de la France, 2009 p.47.

³⁷⁹ Peter VAN DER KNAAP, « Theory-based Evaluation and learning: Possibilities and Challenges », (2004) 10-1 *Evaluation* 16-34.

³⁸⁰ Michael Quinn PATTON, *Utilization-focused evaluation : the new century text*, Thousand Oaks; London; New Delhi, Sage Publications, 1997.

³⁸¹ Il ne faut pas confondre théories d'évaluations et l'évaluation basée sur la théorie qui est une démarche d'évaluation particulière développée dans les années 80s, S. DONALDSON et M. LIPSEY, « Roles for theory in Contemporary Evaluation Practice: Developing Practical Knowledge », dans *Handbook of evaluation*, 2006.

³⁸² Traduction par l'auteure à partir de l'anglais « *The ideal (never achievable) evaluation theory would describe and justify why certain evaluation practices lead to particular kinds of results across situations that evaluators confront. It would (a) clarify the activities, processes, and goals of evaluation; (b) explicate relationships among evaluative activities and the processes and goals they facilitate, and (c) empirically test propositions to identify and address those that conflict with research of other critically appraised knowledge about evaluation* » William R SHADISH, Thomas D COOK et Laura C LEVITON, *op. cit.* p. 31.

³⁸³ Donald T. CAMPBELL, « Reforms as Experiments », *op. cit.*

2. L'utilisation de la connaissance : Comment utiliser la connaissance produite pour améliorer le programme ?
3. Etablir la valeur : Comment construire un jugement global sur un programme pour juger de sa qualité ? Quelle est la notion de bien ?
4. L'évaluation en pratique : comment les évaluateurs devraient orienter leurs pratiques au vu du « monde réel » et du contexte de l'évaluation ?
5. La programmation politique : quels problèmes publics sont résolus par ce dispositif, et peut-il être amélioré ?

Néanmoins, au vu des différentes théories d'évaluation existantes, il semble que peu d'entre elles embrassent l'intégralité de ces problématiques. Comme l'ont souligné M. Alkin et C. Christie, on peut davantage organiser les théories d'évaluation selon leurs spécialisations autour des thématiques de ces problématiques qui relèvent soit, des méthodes, des valeurs ou de l'utilisation des résultats d'évaluation³⁸⁴.

Nous proposons ici, de nous inscrire dans cette littérature en tentant d'élaborer une théorie d'évaluation de type 3 - confrontant méthodologie et utilité publique - de la méthode expérimentale par assignation aléatoire. Nous abordons cette méthode en confrontant la théorie d'évaluation défendue par ses promoteurs à l'épreuve de ses incarnations concrètes. Néanmoins, nous ne souhaitons pas apporter un jugement normatif sur cette méthodologie mais plutôt, par les outils d'analyses de la sociologie de l'action publique, comprendre les logiques d'apprentissages et la nature des interactions qu'elle organise.

Comme nous l'avons vu dans les chapitres précédents, cette méthode d'évaluation est présentée comme le transfert direct d'une méthode scientifique au champ de l'action publique et comme détenant le pouvoir de favoriser l'apprentissage par la preuve, c'est-à-dire par la seule rigueur de sa démonstration. En d'autres termes, les défenseurs de la méthode aléatoire s'inscrivent uniquement dans les débats sur la rigueur de l'administration de la preuve et estiment dès lors que cette dernière suffit à garantir l'utilité et l'utilisation d'une évaluation ayant recours à cette méthode. Nous souhaitons dans ce chapitre, étudier des pratiques de l'utilisation de cette méthode en reprenant en partie les questionnements du courant de recherche sur les *théories d'évaluation* présenté ci-dessus. Nous réunissons ces questionnements autour de deux ensembles de variables d'attention :

- D'une part, nous étudierons les variables liées au contexte de son utilisation, en nous intéressant à l'environnement politique dans lequel elle intervient et les raisons de son utilisation. Cet environnement politique est une variable à étudier tout au long du processus évaluatif puisqu'elle influence aussi bien la mise en œuvre de cette méthode que l'utilisation des résultats (problématiques 5,2,4 évoquées supra) ;
- D'autre part, cette recherche portera sur les éléments liés aux caractéristiques intrinsèques de cette méthode, c'est-à-dire à son protocole de construction de la connaissance, au processus d'apprentissage qu'elle organise, aux interactions particulières qu'elle génère entre les parties prenantes, au type de résultats qu'elle produit et à sa conception de l'action publique et de sa place dans cet environnement (problématique 1,2,3,4 évoquées supra).

Comme nous l'avons précisé dans la première partie de ce chapitre, plusieurs travaux de recherche évaluative ont déjà délimité le périmètre de validité de la démonstration de la

³⁸⁴ Marvin C ALKIN et Christina A CHRISTIE, « An evaluation theory tree revisited », dans Marvin C. ALKIN (dir.), *Evaluation roots: a wider perspective of theorists' views and influences*, 2nd ed, Los Angeles, SAGE Publications, 2004.

causalité proposée par la méthode expérimentale par assignation aléatoire³⁸⁵. Cependant, aucun travail jusqu'à présent n'a abordé l'environnement politique de cette pratique et la relation particulière qu'elle organise entre les acteurs de l'action publique et les chercheurs conduisant des évaluations. Nous estimons que cette méthode scientifique se détache des autres méthodes d'évaluation par la manière dont elle structure l'intervention publique. Ainsi, il paraît indispensable de l'aborder par le prisme de la sociologie de l'action publique, et d'apprécier ses incarnations dans les pratiques afin de comprendre les enjeux de son insertion au sein du champ politique. Ce programme de recherche consistant à dépasser une vision neutre et fonctionnaliste des outils scientifiques mobilisés par l'action publique bénéficie déjà de l'appareillage théorique des *instruments d'action publique*.

Les instruments d'action publique, des institutions d'apprentissage

Nous avons déjà présenté dans notre introduction la pertinence d'étudier la méthode expérimentale comme un instrument d'action publique et donc comme potentiellement une institution sociale. Dans son application médicale, la méthode aléatoire, appelée *essai clinique*, a été promue et étudiée comme une institution sociale dont le but était d'assurer que les médecins utiliseraient le meilleur traitement disponible³⁸⁶. Suivant ce modèle, les économistes, qui ont soutenu le recours à cette méthode dans l'action publique, l'ont présentée comme un instrument favorisant et organisant le développement de la prise en compte de données probantes dans la décision publique³⁸⁷. Si ces auteurs ne l'entendaient pas au sens d'un concept de sociologie de l'action publique, il semble néanmoins que l'on puisse émettre l'hypothèse que cette méthode peut être étudiée comme une institution.

Dans la conception large des néo-institutionnalistes, l'institution sociale est considérée comme une « *procédure organisée et établie* » de « *séquences d'interactions standardisées* »³⁸⁸. Or la méthode expérimentale est la seule méthode d'évaluation qui affecte les individus, organise leur accès à un traitement défini ou à des ressources, et dont la qualité des résultats dépend principalement de la conformité (*compliance*) des individus au protocole établi. Le respect de la répartition des groupes par tirage au sort y est essentiel, autant que le contrôle de l'interaction entre le bénéficiaire et le metteur en œuvre, contrôle qui garantit la transparence de l'analyse des résultats. Nous pouvons donc émettre l'hypothèse qu'elle serait susceptible d'organiser des « *séquences d'interaction standardisées* » tel une institution sociale.

Nous avons vu dans les chapitres précédents que l'expérimentation aléatoire était un instrument porteur de théories économiques. Cette capacité des instruments de mesure statistique à constituer des véhicules de conceptions économiques de la société est un des axes de recherches fondamentaux de la littérature scientifique sur les instruments d'action publique³⁸⁹. Par exemple, C. Radaelli, s'est particulièrement intéressé aux

³⁸⁵ A DEATON, *op. cit.*; Thomas COOK et Michael SCRIVEN, *op. cit.*; Michael SCRIVEN, « Demythologizing Causation and Evidence », dans Stewart I DONALDSON, Christina A CHRISTIE et Melvin M MARK (dir.), *What Counts as Credible Evidence in Applied Research and Evaluation Practice?*, Sage, 2009.

³⁸⁶ Harry MARKS, *op. cit.* p. 99.

³⁸⁷ Esther DUFLO, « Evaluer l'impact des programmes d'aide au développement: le rôle des évaluations par assignation aléatoire », *op. cit.*; Agnès LABROUSSE, « Nouvelle économie du développement et essais cliniques randomisés: une mise en perspective d'un outil de preuve et de gouvernement », *op. cit.* p. 7; Yannick L'HORTY et Pascale PETIT, « L'évaluation aléatoire: un succès qui ne doit rien au hasard », *la vie des idées* (2011).

³⁸⁸ Traduction de l'auteure à partir de l'anglais « *organized, established procedure* » of « *standardized interaction sequences* » Ronald L. JEPPELSON, « Institutions, institutional effects and institutionalism », dans Walter POWELL et Paul DIMAGGIO (dir.), *The New Institutionalism in Organizational Analysis*, Chicago: University of Chicago Press, 1991, p. 143-163.

³⁸⁹ Alain DESROSNIÈRES, « La statistique, outil de gouvernement et outil de preuve », dans *L'argument statistique*, *op. cit.*; Patrick LE GALÈS et Pierre LASCOUMES, « Introduction : Understanding Public Policy through its Instruments - From the Nature of

Regulatory impact analysis (analyses d'impact ex ante), autres instruments de l'*evidence-based policy* employés par l'Union Européenne³⁹⁰. Il met à jour dans cette analyse la propriété des instruments de l'*evidence-based policy* à être aussi porteur d'un modèle d'apprentissage³⁹¹. Cette dimension d'apprentissage apparaît comme essentielle dans notre étude.

L'apprentissage politique a été l'objet d'une variété de travaux. Cependant ces derniers ne se retrouvent pas unifiés autour d'un noyau théorique ou méthodologique³⁹². À leur lecture, deux catégories de réflexions se détachent, une de type taxinomique, s'intéressant à élaborer des typologies d'apprentissages, et une autre attachée aux variables exerçant une influence sur l'apprentissage politique.

Ces travaux portent sur une variété de types d'apprentissages politiques possibles, allant du transfert d'un instrument à la mode, à l'apprentissage réflexif. H. Heclo décrit le processus de construction des politiques publiques comme la résolution d'un casse-tête (*puzzling*). Ce casse-tête est compris comme un phénomène d'apprentissage, de recherche de solutions en mettant donc en valeur l'importance de la dimension cognitive dans les politiques publiques. Pour cet auteur, l'administration est dans un apprentissage permanent dans la conception des politiques publiques – ce qu'il appelle le « *Political learning* » (apprentissage politique)³⁹³. Ce concept désigne l'apprentissage de façon très générale, puisqu'il comprend les apprentissages portant autant sur le processus politique que sur les politiques publiques. L. Etheredge développe le concept voisin d'apprentissage gouvernemental (*government learning*) qu'il définit comme l'augmentation des connaissances du gouvernement dans un souci d'efficacité³⁹⁴. Cet apprentissage d'acteurs influents ou d'institutions formelles porte sur l'organisation du gouvernement. Son objet étant l'efficacité organisationnelle, il n'est donc pas pertinent pour notre étude. Les travaux de P. Sabatier positionnent le « *policy oriented learning* » (l'apprentissage à propos de la politique publique) au cœur de la détermination du changement et de l'innovation dans le domaine de l'intervention publique³⁹⁵. Cet apprentissage se fait par un sous-système politique, une coalition de cause, constituée d'acteurs étatiques ou non, unis autour d'une insatisfaction quant aux arrangements institutionnels actuels, que ce soit sur le plan des connaissances mobilisées ou celui du mode d'interaction. L'apprentissage porte sur la compréhension d'une relation causale ou la capacité à répondre à un problème. Il serait donc susceptible d'être pertinent pour notre étude. Enfin, Richard Rose développe le concept de « *lesson drawing* » (littéralement le fait de *tirer* des leçons de quelque chose ou quelqu'un). L'objet de l'apprentissage, par ce concept, concerne les politiques publiques insatisfaisantes, c'est un apprentissage qui passe soit par les compréhensions des erreurs passées, soit par l'enseignement d'expériences étrangères. Ainsi, R. Rose souligne que l'évaluation et le « *lesson drawing* » sont liés de façon inextricable puisque chaque leçon suppose un jugement sur une action passée³⁹⁶. Enfin, quant aux effets du changement, R. Rose

Instruments to the Sociology of Public Policy Instrumentation », (2007) 20-1 *Governance: An International Journal of Policy, Administration, and Institutions* pp1-21.

³⁹⁰ Paola COLETTI et Claudio M. RADAELLI, « Economic Rationales, Learning, and Regulatory Policy Instruments », (2013) 91-4 *Public Administration* 1056-1070.

³⁹¹ *Ibid* p. 1062.

³⁹² M EVANS et J DAVIES, « Understanding policy transfer: a multi-level, multi-disciplinary perspective », *Public Administration* 1999.

³⁹³ Hugh HECLO, *Modern social politics in Britain and Sweden : from relief to income maintenance*, New Haven [u.a.], Yale Univ. Press, 1975.

³⁹⁴ Lloyd S. ETHEREDGE, « Government Learning: An Overview », dans *The Handbook of Political Behavior* vol. 2, Plenum Press, New York, 1981, p. 73-161.

³⁹⁵ C. BENNETT et M. HOWLETT, « The lessons of learning: Reconciling theories of policy learning and policy change », *Policy Sciences* 1992.25.275- 294, p. 82.

³⁹⁶ Richard ROSE, « What is Lesson-drawing? », *Journal of Public Policy* 1991.11, p.19.

s'intéresse surtout au changement d'instruments ou de programmes.

Notre objet semble à la fois correspondre au *policy oriented learning* mais dont la coalition de cause n'est pas constituée autour de la promotion d'une politique publique, mais d'une démarche d'élaboration de politiques publiques, la méthode expérimentale. Ensuite, le concept de *lesson drawing* semble pour l'instant trop vaste, puisqu'il comprend l'apprentissage issu de ses propres erreurs ou des expériences étrangères.

J. Bennett et M. Howlett précisent que l'opérationnalisation de cette classification passe par un travail repositionnant le choix de ce qui est appris et de ce dont on se souvient dans un contexte d'intérêts politiques et de pouvoir³⁹⁷. Pour reprendre le vocabulaire de H. Heclo, il faut réintégrer du « *powering* » c'est-à-dire l'étude de la lutte entre intérêts différents pour l'appropriation des ressources, pour la compréhension du phénomène de « *puzzling* », au sens où le questionnement autour de la résolution de problème est fondamentalement lié en politique aux enjeux de pouvoir. La littérature américaine de ces vingt dernières années témoigne d'une volonté de rendre compte de l'importance des enjeux politiques dans les processus d'apprentissage. Dans son article « *Political conflict and lesson drawing* », D. Robertson relève la tendance à poser les adoptions de politiques publiques comme des « *vérités neutres politiquement* » tandis que, selon lui, « *les adversaires utilisent ces leçons comme des armes politiques* »³⁹⁸. Les déterminants de l'apprentissage politique sont donc à chercher dans des variables liées aux enjeux politiques ou de pouvoir propres aux décideurs publics ou à l'intervention publique. S'intéresser à l'apprentissage requiert alors un travail préalable d'identification des intérêts des parties prenantes, et de ceux qui seront le mieux servis par ce processus d'apprentissage³⁹⁹. Ces travaux suggèrent que l'apprentissage à partir d'une preuve serait uniquement possible si cette dernière sert les intérêts des décideurs publics.

Enfin, la littérature de l'apprentissage organisationnel propose une autre catégorisation d'apprentissage autour de la nature du processus cognitif à l'œuvre. C. Argyris et D. Schön distinguent deux types d'apprentissage, l'apprentissage à simple boucle (*single-loop learning*) et l'apprentissage à double boucles (*double-loop learning*)⁴⁰⁰. Dans le cadre d'un apprentissage à simple boucle, les acteurs modifient leur action quand ils constatent une différence entre les résultats souhaités et les résultats effectifs. C'est le procédé régulier d'identification d'un problème, et de sa correction à la lumière de ce qui a été appris. Dans le cadre d'un apprentissage à deux boucles, la première étape de résolution de problème est dépassée pour questionner en profondeur les présupposés, hypothèses, pratiques, valeurs, qui ont été à l'origine du problème dans le but d'agir pour modifier celles-ci afin qu'elles ne provoquent plus ce problème. En d'autres termes si le premier type d'apprentissage cherche une solution pour répondre à un problème, le second type d'apprentissage propose une solution pour supprimer les causes de ce problème⁴⁰¹. L'apprentissage politique semble surtout se situer dans de l'apprentissage de premier ordre.

Dans un effort de clarification et d'unification des travaux sur l'apprentissage politique, C. Dunlop et C. Radaelli ont récemment proposé une typologie d'apprentissages

³⁹⁷ C. BENNETT et M. HOWLETT, *op. cit.* p. 289.

³⁹⁸ Traduction de l'auteure à partir de l'anglais « *policy lessons from abroad often are put forward as politically neutral truths. Beneath this [...] adversaries are just as often using such lessons as political weapons* » D. ROBERTSON, « Political Conflict and Lesson Drawing », (1991) 11-1 *Journal of Public Policy* 55-91, p.55.

³⁹⁹ D. DOLOWITZ et D. MARSH, « Who Learns What from Whom : a Review of the Policy Transfer Literature », (1996) 44-2 *Political Studies* 342-357.

⁴⁰⁰ CHRIS ARGYRIS et DONALD A SCHÖN, *A theory of action perspective*, Reading, Mass [u.a.], Addison-Wesley Pub. Co, 1978.

⁴⁰¹ MICHAEL QUINN PATTON, *Developmental evaluation : applying complexity concepts to enhance innovation and use*, New York, Guilford Press, 2010, p. 11.

politiques davantage opérationnelle pour notre objet. Embrassant une acception large de l'apprentissage comme étant « *l'actualisation de croyances basées sur des expériences vécues ou observées, l'analyse ou l'interaction sociale* », ces auteurs proposent quatre genres (au sens de subdivision biologique) d'apprentissages politiques⁴⁰². Ces genres sont organisés autour d'une matrice classant, d'une part selon le degré d'incertitude lié au problème et la facilité à le résoudre (*problem tractability*), et d'autre part selon le degré de certification requis et reconnu pour les formateurs (*certification of actors*).

Figure 4: Les quatre genres identifiés dans la littérature sur l'apprentissage politique⁴⁰³

		PROBLEM TRACTABILITY	
		LOW	HIGH
CERTIFICATION OF ACTORS	LOW	Reflexive Learning	Learning through Bargaining
	HIGH	Epistemic Learning	Learning in the Shadow of Hierarchy

Nous ne reviendrons pas ici sur le détail de ces quatre genres, mais nous pouvons dès lors établir que la méthode aléatoire et l'apprentissage qu'elle propose s'inscrivent dans l'apprentissage épistémique (*Epistemic learning*). Pour ce « genre », la connaissance est restreinte à un ensemble d'experts limité (haute certification des acteurs) pour recentrer la discussion autour d'enjeux techniques (problème difficile à résoudre, soit *low tractability*) de politiques publiques. La littérature s'intéressant à ce genre d'apprentissage se concentre sur l'impact de la rationalité sur le changement de politiques publiques notamment via des instruments institutionnels comme les instruments d'action publique⁴⁰⁴.

Afin d'affiner cette typologie, ces auteurs s'intéressent aussi au processus d'apprentissage politique, et non pas uniquement à l'utilisation des résultats des travaux scientifiques. Dimension qui, comme nous l'avons évoqué plus haut, a fait l'objet d'une profusion de travaux dans la littérature de la recherche évaluative. S'appuyant sur les travaux de D. Mocker et G. Spear, C. Dunlop et C. Radaelli proposent au sein de chaque genre d'apprentissage politique quatre « espèces » organisées selon le degré de contrôle des apprenants, d'une part sur le but et les objectifs de l'apprentissage, et d'autre part sur les moyens et le contenu de l'apprentissage.

⁴⁰² Traduction de l'auteure à partir de l'anglais " *updating of beliefs based on lived or witnessed experiences, analysis or social interaction*" Claire A. DUNLOP et Claudio M. RADAELLI, « Systematising Policy Learning: From Monolith to Dimensions », (2013) 61-3 *Political Studies* 599-619.

⁴⁰³ Claire A. DUNLOP et Claudio M. RADAELLI, *op. cit.*

⁴⁰⁴ *Ibid* p.600

Figure 5: Les quatre espèces d'apprentissages du genre « apprentissage épistémique »⁴⁰⁵

		LEARNERS' CONTROL OVER LEARNING OBJECTIVES / ENDS	
		HIGH	LOW
LEARNERS' CONTROL OVER LEARNING CONTENT / MEANS	HIGH	CONTRIBUTOR	PRODUCER OF STANDARDS
	LOW	FACILITATOR	<u>TEACHER</u>

Quatre espèces d'apprentissages apparaissent au sein du genre « apprentissage épistémique » caractérisant le statut du formateur organisant l'apprentissage :

- L'enseignant (*teacher*). Cette relation est l'idéal type de la relation enseignant et apprenant. L'enseignant contrôle la fin et les moyens de l'apprentissage, et contribue à la définition des intérêts des acteurs de l'action publique.
- Le contributeur (*contributor*). Le formateur est ici considéré comme un intervenant parmi d'autres et les preuves qu'il apporte peuvent être contestées par les acteurs de l'action publique. Ces derniers contrôlent leur apprentissage, leurs interprétations et objectifs.
- Le facilitateur (*facilitator*). Les formateurs ne participent pas à la définition des objectifs d'apprentissage. Ce sont des agents contrôlés par les acteurs de l'action publique qui contribuent à faciliter le processus d'apprentissage.
- Le producteur de standards (*producer of standards*). Les formateurs établissent les buts de l'action publique auxquels doivent répondre les acteurs de l'action publique, mais ce sont ces derniers qui ont à leur charge d'apporter les preuves qu'ils s'inscrivent dans ces standards. Comme par exemple face à des standards internationaux fixés par des experts.

Les auteurs précisent néanmoins que ces types d'apprentissages ne sont pas figés selon les problèmes publics ou les experts. Un même problème ou les mêmes experts peuvent passer d'une case à l'autre dans le temps⁴⁰⁶.

Ce travail ouvre un nouvel agenda de recherche sur l'identification du modèle d'apprentissage véhiculé par la méthode expérimentale par assignation aléatoire et de sa pertinence pour le champ de l'action publique. Nous chercherons à identifier à laquelle de ces « espèces » d'apprentissage correspondent les incarnations de pratiques de la méthode aléatoire, en étant attentif aux possibilités de glissement d'une catégorie à l'autre, soit d'une expérimentation à l'autre, soit dans le temps au sein d'une même expérimentation. Néanmoins, nous pouvons déjà faire l'hypothèse que la méthode expérimentale constitue un objet particulier difficilement classifiable du fait de la poursuite de deux intérêts d'apprentissages, celui du chercheur et celui des acteurs de l'action publique.

⁴⁰⁵ Claire A. DUNLOP et Claudio M. RADAELLI, *op. cit.* p. 609

⁴⁰⁶ Claire A. DUNLOP et Claudio M. RADAELLI, *op. cit.* p. 615.

Avant d'identifier l'espèce ou les espèces d'apprentissages relatives à cette méthode, il est nécessaire d'outiller notre réflexion sur la particularité de cette relation entre deux champs aux pratiques opposées, organisée par la méthode aléatoire.

2.2.1.2. Analyser les séquences d'interactions organisées par l'expérimentation randomisée : les sites institutionnels de la relation clinique

Le chapitre 3 mobilise à son tour les cadres théoriques du *travail institutionnel* ainsi que des *champs d'action stratégique*⁴⁰⁷. Néanmoins, nous souhaitons maintenant éprouver ces cadres à un niveau microsociologique, en d'autres termes en les appliquant au niveau de l'individu ou de petits groupes. L'échelle d'analyse devient ici l'expérimentation et ses acteurs. La *théorie générale des champs* est mobilisée afin de rendre compte et d'expliquer les motivations des acteurs pour s'engager dans les expérimentations évaluées par la méthode aléatoire contrôlée. Le concept de *travail institutionnel* nous permet d'analyser les confrontations de pratiques au sein de l'institution temporaire de *relation clinique* organisée par l'expérimentation par assignation aléatoire.

La relation clinique, ou les enjeux d'un double apprentissage

L'évaluation d'expérimentations par la méthode expérimentale par assignation aléatoire a été présentée dans le milieu scientifique et dans celui de l'action publique comme organisant une relation de « gagnant/gagnant ». D'une part elle permettait aux premiers d'avoir accès à de nouvelles données, d'autre part aux seconds, de disposer de la preuve indéniable de l'impact de leur programme. Cette méthode devenait ainsi un moyen d'instituer une *relation clinique* de double apprentissage entre les scientifiques et les acteurs de l'action publique, au sens où la même recherche devait trouver autant une valeur académique pour les chercheurs qu'une valeur opérationnelle pour les décideurs publics⁴⁰⁸. R. Ripley, qui a forgé ce concept, comprend le terme clinique comme « *extrêmement objectif et réaliste* » « *traitant de ou s'appuyant sur de véritables observations et traitements de maladies sur des patients plutôt que des expérimentations artificielles ou des théories* »⁴⁰⁹. Cette démarche clinique, souhaitée, s'apparente donc à la démarche médicale de praticien clinique, lequel conduit des recherches sur ses patients en même temps qu'il les soigne⁴¹⁰. R. Ripley identifie quatre composantes de cette relation⁴¹¹:

- une coentreprise (*joint venture*) où chacune des parties apprend et enseigne à l'autre ;
- un esprit scientifique, se manifestant par la recherche de l'objectivité, le réalisme et la rigueur analytique ;
- un apprentissage lié à une observation empirique ;
- et qui a pour but de résoudre des problèmes pratiques de court terme et d'établir les bases pour une théorie future.

Néanmoins, s'appuyant sur les travaux de I. L. Horowitz, cet auteur reconnaît qu'une telle relation souffre des divergences d'orientations et de procédures du champ politique et scientifique, puisque si les membres du second recherchent la compréhension

⁴⁰⁷ Thomas B. LAWRENCE et Roy SUDDABY, « Institutions and Institutional Work », *op. cit.*

⁴⁰⁸ Randall B. RIPLEY, *op. cit.* p. 206.

⁴⁰⁹ L'auteur reprend ici la définition du Unabridged Random House Dictionary of the English Language Randall B. RIPLEY, *Ibid.* p. 208.

⁴¹⁰ Evert VEDUNG, « Four waves of evaluation diffusion », (2010) 3-16 *Evaluation* 263-277, p.274.

⁴¹¹ Randall B. RIPLEY, *op. cit.* p. 208.

(*understanding*), les acteurs de l'action publique recherchent un guide pratique pour l'action⁴¹². Il identifie quatre points critiques d'interactions entre ces deux parties au cours de cette *relation clinique* temporaire⁴¹³ :

- pendant la conception de la recherche, où chacun s'assure que cette dernière répondra aux intérêts de chacun ;
- lors de la planification du calendrier du projet, afin d'assurer que l'analyse soit produite à temps pour son utilisation ;
- lors de la production des premiers rapports intermédiaires, ce qui peut amener à des négociations à mi-parcours ;
- et enfin la dernière interaction sur la signification des résultats et les implications ou recommandations pour l'action.

Ces interactions sont identifiées comme autant de lieux de négociations entre les parties prenantes de la *relation clinique*. En effet, si on met en perspective ce concept avec les travaux évoqués précédemment, cela implique que dans une *relation clinique* la recherche d'apprentissage est autant présente chez l'apprenant que chez le formateur. Le formateur/chercheur n'est pas au service de l'apprentissage du porteur de projet. L'apprentissage de ce dernier n'est qu'une conséquence du propre apprentissage du chercheur. Cet aspect dédouble les enjeux de l'apprentissage politique puisqu'on peut émettre l'hypothèse que des processus de *puzzling* (recherche de solution) et de *powering* (lutte de pouvoir pour le contrôle des ressources) sont présents du côté des deux parties de la relation d'apprentissage, mais qu'il faut néanmoins que ces dernières convergent vers des objectifs communs ainsi qu'une méthodologie d'apprentissage commune.

Les champs d'actions stratégiques et le travail institutionnel pour comprendre les enjeux de la relation clinique

Ce travail propose une conception théorique s'appuyant sur ce modèle de *relation clinique* et l'enrichissant par les concepts de *champ d'action stratégique* et de *travail institutionnel*. La notion de *relation clinique* met en valeur trois caractéristiques de ce type particulier d'institution d'apprentissage.

D'abord, elle souligne la relation d'interdépendance entre les deux parties en interactions dans l'expérimentation⁴¹⁴. Le porteur du projet expérimental ne peut prétendre conduire une expérimentation aléatoire sans un évaluateur, et en retour le chercheur souhaitant conduire des travaux avec cette méthode doit avoir l'autorisation du porteur de projet pour mettre en œuvre son protocole. On peut donc considérer l'expérimentation comme une *structure* d'interaction au sens de l'épistémologie réaliste critique⁴¹⁵. Les structures y sont comprises comme des relations « d'objets » interreliés qui définissent leur identité les uns par rapport aux autres. En d'autres termes, dans le cadre d'une expérimentation par assignation aléatoire, le rôle de porteur de projet expérimental ne se définit que vis-à-vis d'un évaluateur expérimental et réciproquement⁴¹⁶. Cette structure entre en mouvement par des mécanismes qui génèrent

⁴¹² Irving Louis HOROWITZ, « Social science mandarins: Policymaking as a political formula », (1970) 1-3 *Policy Science* 339-360.

⁴¹³ Randall B. RIPLEY, *op. cit.* p. 215.

⁴¹⁴ Nous verrons plus loin, que lorsque le financeur n'est pas le porteur de projet, on assiste davantage à un triangle d'interdépendance.

⁴¹⁵ Eric W. K. TSANG et Kai-Man KWAN, « Replication and Theory Development in Organizational Science: A Critical Realist Perspective », *op. cit.*

⁴¹⁶ Nous entendons par le concept de rôle « une typification de conduites socialement objectivées » passant par un processus de socialisation comprise comme « l'intériorisation de sous-mondes institutionnels spécialisés (...) l'acquisition de savoirs

des actions entre ces éléments selon des conditions intrinsèques propres à chaque mécanisme, et des conditions extrinsèques variant selon les contextes d'applications⁴¹⁷. C'est l'identification de ces mécanismes et leurs conditions intrinsèques et extrinsèques d'activation qui seront au cœur de notre étude.

Cet objet nous amène à la deuxième caractéristique de ce type particulier d'institution. La méthode expérimentale par assignation aléatoire vise à structurer les interactions en son sein, uniquement sur une période de temps limité, celui de l'expérimentation. Ainsi, les rôles incarnés au sein de cette structure ne le sont que pour une période donnée. Néanmoins, cette institution nécessite une conformité stricte de ses acteurs aux règles de cette dernière, comme le contrôle scientifique du « traitement », de son administration aléatoire, ainsi que l'attente de la production des résultats scientifiques. La particularité de cette institution est que les rôles qu'elle organise et ses normes nécessitent un apprentissage rapide des acteurs qui n'ont pas bénéficié d'une socialisation préalable pour se les approprier. Nous serons ici attentifs à un déséquilibre de socialisation des rôles entre les parties prenantes de l'expérimentation. En effet, cette institution émanant du champ de l'économie expérimentale et adoptant les règles de production scientifique semble plus accessible aux économistes s'engageant comme évaluateurs qu'aux porteurs de projets ou acteurs de la mise en œuvre. D'autant plus que les premiers participent à plusieurs expérimentations et que chacune participe à l'apprentissage du rôle d'évaluateur expérimental.

Ainsi, cette institution est particulièrement sujette au *travail institutionnel*, c'est-à-dire à l'« *action intentionnelle des acteurs ou des organisations visant à créer, maintenir ou déstabiliser les institutions* »⁴¹⁸. Elle concentre sur un temps limité à la fois des actions de création, de maintien et de déstabilisation émanant des parties prenantes de l'expérimentation. Nous identifions trois séquences d'interactions et de négociations distinctes ayant des enjeux propres au sein de l'expérimentation aléatoire. Une séquence prévue par R. Ripley est absente, celle de la planification du calendrier du projet assurant la production des résultats à temps pour leur utilisation. Nous reviendrons sur l'impact de l'absence de cette étape lors de notre étude de l'utilisation des résultats. Notre découpage est donc le suivant :

1. L'étape de conception et planification de la recherche, où les deux parties doivent s'assurer que la recherche répond aux intérêts de chacune des parties
2. La négociation de mi-parcours : qui correspond au suivi du protocole, aux ajustements liés à la mise en œuvre, ou aux premiers résultats produits pouvant susciter une renégociation des termes de la relation
3. L'interaction finale, le moment de la production des résultats et de leur interprétation et utilisation par les parties prenantes.

Cette dimension appelle à préciser la dernière caractéristique de ce type particulier d'institution, qui réside dans l'opposition entre les pratiques des champs qu'elle met en relation. La *relation clinique* organise une institution temporaire entre le *champ d'action stratégique* de l'économie appliquée expérimentale et celui de l'action publique. Bien

spécifiques et de rôles directement ou indirectement enracinés dans la division du travail » Peter L BERGER et Thomas LUCKMANN, *La construction sociale de la réalité*, Paris, Armand Colin, 2006, p. 103.

⁴¹⁷ Marie-José AVENIER et Catherine THOMAS, « Mixer Quali et Quanti pour quoi faire? Méthodologie sans épistémologie n'est que ruine de la réflexion! », *Cahier de Recherche CERAG* 2011, p.8.

⁴¹⁸ L'auteur reprend ici la traduction de l'anglais : « *the purposive action of individuals and organizations aimed at creating, maintaining and disrupting institutions* » proposée dans, Ben Slimane, Karim, et Bernard Leca. *op. cit.*

sûr, le *champ d'action stratégique* de l'action publique est lui-même composé de sous-champs distincts selon les secteurs d'activité concernés et la nature des structures la conduisant. L'étude des enjeux de ces *champs d'action stratégiques* respectifs nous informe sur les motivations des différentes parties prenantes à accepter cette institution. Elle nous donne aussi les clefs pour analyser la nature du travail institutionnel de chacun, lequel témoigne de la difficulté d'inscrire ces *pratiques* – appartenant à une logique de champ plus globale et s'appuyant sur un ensemble spécifique de ressources, de compétences et de perceptions du monde – dans un nouveau rôle institutionnel, ouvrant ainsi sur l'étude de la confrontation d'un travail de maintien et de déstabilisation institutionnels, ou de l'atteinte d'un consensus entre les différentes parties. Cette dimension, qui envisage l'institution comme une structure malléable et évolutive, ouvre sur l'étude du processus d'apprentissage en tant que tel ainsi que sur une interrogation quant à l'évolution au cours de cette relation des modalités d'apprentissages et de ses objectifs.

Notre conception interactionnelle de l'institution se rapproche donc, comme l'a déjà souligné O. Nay, de la notion de *configuration* de N. Elias, au sens où l'application des règles de l'institution sont soumises à un processus relationnel et sont inscrites dans des situations concrètes d'échange, ouvrant à l'identification de plusieurs *modèles de jeux* d'acteurs⁴¹⁹. O. Nay propose alors le concept de *sites institutionnels* qui soutient l'idée qu'une institution n'est pas une structure composée de règles figées dans le temps et l'espace, mais s'incarne de manière différenciée selon les dispositions des parties prenantes à entrer dans les jeux⁴²⁰. Il définit les sites institutionnels comme « *des espaces sociaux identifiés par trois critères : un ensemble d'acteurs interdépendants (configuration) ; l'existence d'interactions menées en continu (dynamique de jeux) ; et enfin la présence de règles institutionnalisées exprimées dans un ensemble de manières d'agir et de penser (niveau microsociologique) et de logiques sociales (niveau macrosociologiques) pesant sur l'ensemble des comportements et des usages des acteurs en situation de co-présence* »⁴²¹. Nous reprendrons ici le programme de recherche proposé par cet auteur, dont nous estimons qu'il correspond à la stratégie de recherche et à l'épistémologie du réalisme critique que nous avons adoptées, lesquelles cherchent à « *rendre compte de certaines logiques sociales spécifiques qui sont propres à l'institution politique (et scientifique pour notre objet) observée, et d'autre part à saisir la pluralité des usages et des représentations que ces logiques induisent dans des contextes d'interaction particuliers mettant aux prises des acteurs porteurs de rationalités et d'exigences différenciées.* ». Ainsi, nous proposons de comprendre l'institution d'apprentissage de l'expérimentation aléatoire par l'étude de ses variations d'incarnation, nous permettant aussi de mettre à jour ses enjeux et logiques spécifiques.

Opérationnalisation de la construction théorique

Nous rappelons que ce chapitre porte sur deux questionnements. Il nourrit d'une part notre questionnement général sur l'impact de cette méthode sur la reconfiguration des interactions entre les sciences sociales et l'action publique et d'autre part questionne l'adaptabilité de cette méthode aux modes d'apprentissages des acteurs de l'action publique.

⁴¹⁹ Olivier NAY, *La région, une institution: la représentation, le pouvoir et la règle dans l'espace régional*, op. cit. p. 17.

⁴²⁰ *Ibid.* p. 18.

⁴²¹ Olivier NAY, op. cit. p. 20.

Nous émettons ici quatre hypothèses répondant à ces deux questions de recherche. Ces hypothèses sont issues de notre réflexion théorique ainsi que de nos premières observations empiriques.

La conception de la méthode aléatoire comme une *relation clinique* aux prises d'un *travail institutionnel*, nous invite à nous intéresser dans un premier temps à identifier l'impact particulier de cette méthode sur ses parties prenantes. Nous émettons ici l'hypothèse que :

Hypothèse 3.1 : La pratique de cette méthode organise de nouveaux types d'interactions entre les acteurs de l'action publique et les milieux scientifiques, mais ces interactions sont variables d'un site institutionnel à l'autre selon l'équilibre atteint entre les différentes activités de *travail institutionnel* en leur sein.

Cette hypothèse ouvre un autre questionnement sur les facteurs influençant ces différents types de *travail institutionnel* ainsi que leur pondération. Nous émettons ici la deuxième hypothèse que :

Hypothèse 3.2 : Les enjeux liés au *powering* au sein de l'expérimentation aléatoire conditionnent les activités de *puzzling* pour chaque partie investie dans une expérimentation aléatoire et s'inscrivent dans les logiques de champs d'action stratégiques d'appartenance des parties prenantes.

Une deuxième dimension de notre travail consiste à s'interroger sur l'apprentissage organisé par la méthode aléatoire et notamment sur son adaptabilité aux champs de l'action publique. Nous appuyant sur les travaux de théorie d'évaluation ainsi que sur ceux portant sur les instruments de l'action publique, nous distinguons le processus d'apprentissage en tant que tel de la production scientifique issue de ce processus. Au vu des précédents travaux en recherche évaluative ainsi qu'en action publique, nous émettons les deux hypothèses suivantes :

Hypothèse 3.3 : Le processus de production de connaissance de la méthode aléatoire constitue un obstacle à l'apprentissage des acteurs de l'action publique pendant la conduite de l'expérimentation

Hypothèse 3.4 : L'apprentissage recherché par les chercheurs dans les résultats produits par la méthode aléatoire ne correspond pas aux types de connaissances nécessaires à l'apprentissage des acteurs de l'action publique.

Avant de présenter les réponses apportées par notre recherche empirique à ces hypothèses, il nous faut d'abord revenir sur l'appareil méthodologique empirique particulier à ce chapitre.

2.2.2. Démarche empirique du chapitre 3

Il est ici nécessaire de revenir sur l'appareillage empirique élaboré pour collecter ces données, puisque ce dernier a été spécifique au chapitre 3. Nous défendons dans ce développement, la pertinence de la mobilisation de *l'étude de cas* comme outil le mieux à même de rendre compte de phénomènes contemporains dont les éléments propres à leurs structures intrinsèques et aux contextes dans lesquels ils s'incarnent sont particulièrement liés⁴²². Nous précisons ensuite la stratégie d'analyse croisée des cas, laquelle implique une comparaison diachronique et synchronique. Enfin nous revenons sur les modalités de généralisation et de contrôles des biais subjectifs, par la présentation du protocole de collecte de données et la triangulation de ces dernières.

2.2.2.1. Les études de cas pour une analyse compréhensive des interactions et de leur périphérie

Nous proposons dans ce développement un retour sur les qualités de la conduite d'études de cas pour rendre compte de notre objet. Cette approche, qui peut avoir l'inconvénient de se restreindre à l'étude d'une unité d'analyse aux dépens de son environnement, est complétée par une démarche « emboîtée » permettant de varier les unités d'analyse et ainsi d'appréhender les dynamiques de champs environnant l'expérimentation aléatoire. Se posent alors des questions de possibilités de montée en généralité à partir d'études de cas particulières, à laquelle nous répondrons par une comparaison par induction analytique des cas entre eux afin de mettre à jour les régularités ainsi que les contingences de nos résultats.

Une stratégie de recherche est déterminée à partir de la nature de la question de recherche, le contrôle que peut avoir le chercheur sur les événements observés ainsi que leur temporalité. Les études de cas sont particulièrement appropriées pour les questions de recherche explicatives et/ou descriptives, s'intéressant au « pourquoi » ou au « comment », lorsque le chercheur ne peut manipuler les événements, et lorsque ceux-ci sont contemporains⁴²³. L'étude de cas a longtemps été reléguée à la phase exploratoire d'une recherche, elle était perçue comme génératrice de données uniquement descriptives, et on lui préférait l'expérimentation pour les dimensions explicatives d'un travail de recherche⁴²⁴. Néanmoins, de plus en plus, elle est utilisée pour apporter une compréhension détaillée d'un phénomène dans le domaine des sciences sociales, particulièrement dans les domaines de la science politique, de la sociologie ou de l'anthropologie⁴²⁵. Selon John Gerring, l'étude de cas se trouve dans une situation paradoxale puisque la plupart des connaissances accumulées en sciences sociales, particulièrement en science politique, sont issues d'études de cas, néanmoins cette démarche reste souvent ignorée méthodologiquement et peu reconnue. Nous proposons ici de revenir sur notre emploi de l'étude de cas. Nous nous sommes autant appuyée sur des travaux méthodologiques de sociologie que d'évaluation pour élaborer cette stratégie de recherche car, comme nous l'avons expliqué par ailleurs, il nous semble que ces démarches se renforcent mutuellement pour établir la rigueur de ce cheminement⁴²⁶.

⁴²² Robert K YIN, *Case study research : design and methods*, Thousand Oaks, Calif., Sage Publications, 2003, p. 13.

⁴²³ Robert K YIN, *op. cit.* p. 3.

⁴²⁴ *Ibid.*

⁴²⁵ John GERRING, « What Is a Case Study and What Is It Good for? », (2004) 98-2 *The American Political Science Review* 341-354, p.341.

⁴²⁶ A ce sujet voir notre article "[Les défis de l'emploi de l'étude de cas en évaluation](#)", avec Amandine Grégot, politiques sociales et familiales, n°110, décembre 2012, pp33-43

Le cas, *casus* en latin, désigne une configuration originale d'un agencement de faits qui fait rupture par rapport aux configurations habituelles, et dont la singularité reste inexpliquée⁴²⁷. On retrouve cette même signification quand on se réfère à un cas clinique, ou juridique. En effet, l'étude de cas en sociologie émane d'une situation paradoxale, une situation qui fait problème, qui dévie des cadres théoriques explicatifs usuels, mais dont l'étude appelle la résolution de cette exception par la définition de cadres nouveaux de raisonnement. D'ailleurs, un événement historique en lui-même ne constitue pas un cas à proprement parler, c'est son articulation avec une théorie, un questionnement, qui en fait un cas susceptible de faire l'objet d'une étude⁴²⁸. Une étude classique devient un cas seulement lorsqu'elle est intégrée de façon pertinente à une théorie explicative générale. D'ailleurs, comme le souligne S. Lipset, contrairement aux croyances générales, le but premier de l'étude de cas n'est pas la particularisation, la compréhension d'un phénomène dans sa singularité, mais bien la généralisation rendue possible par l'inscription de chaque cas dans une théorie⁴²⁹. Selon cette approche sociologique du cas, notre cas serait la méthode expérimentale et les termes dans lesquels elle questionne les configurations de l'interaction entre la science et l'action publique. Néanmoins, ce cas nécessite d'être divisé en sous-entités afin de cadrer l'étude méthodique de ce phénomène et de l'étudier selon les préceptes de la méthodologie d'étude de cas.

L'étude de cas se détache des autres méthodes de collecte de données par l'importance qu'elle accorde à la compréhension approfondie du cas étudié. Aux prémices de la méthodologie de l'étude de cas, on retrouve l'ethnométhodologie et l'observation participante de B. Malinowsky qui furent ensuite formalisées pour l'étude monographique de milieux urbains par l'école de Chicago à partir des années 1920. Cette méthode de recherche exigeait du chercheur qu'il s'insère dans la communauté étudiée, à « visage couvert ou découvert », pour en proposer une étude « en contexte ». Si aujourd'hui de nombreux chercheurs pratiquant l'étude de cas ne s'immergent pas en continu et pour une longue durée au contact de leur objet d'étude, la recherche d'une connaissance approfondie et contextualisée du phénomène étudié demeure une condition préalable. Cette proximité est souvent apportée par des outils de collectes de données qualitatives comme l'observation, participante ou non, l'entretien, individuel ou de groupe, l'analyse de données secondaires et particulièrement des documents de littérature grise (documents administratifs, comptes rendus de réunion) ou documents de presse. Cependant, l'étude de cas n'a pas uniquement recours à des outils de collecte de données qualitatives, le chercheur peut aussi recourir à des outils de collecte de données quantitatives si ces dernières peuvent apporter un éclairage sur certaines dimensions du cas étudié. Comme le souligne S. Crasnow, la pratique d'étude de cas en science politique témoigne plutôt de l'utilisation d'une combinaison de méthodes qualitatives et quantitatives⁴³⁰. En croisant les visions communes de plusieurs auteurs, l'étude de cas peut être définie comme *l'étude empirique intensive d'une unité singulière, contemporaine, prise dans son contexte de tous les jours, et ayant pour but la compréhension d'un plus grand ensemble d'unités comparables*⁴³¹.

⁴²⁷ Jean-Claude PASSERON et Jacques REVEL, *Penser par cas*, Paris, Éditions de l'École des Hautes Études en Sciences Sociales, 2005, p. 16.

⁴²⁸ Michel WIEVIORKA, « Case studies: history or sociology? », dans Charles C RAGIN et Howard Saul BECKER, *What is a case? : exploring the foundations of social inquiry*, Cambridge [England]; New York, NY, USA, Cambridge University Press, 1992.

⁴²⁹ Lipset in Robert K YIN, *op. cit.*

⁴³⁰ Sharon CRASNOW, *Evidence for Use: The Role of Case Studies in Political Science Research*, Minnesota, SPSP 2009: Society for Philosophy of Science in Practice, 18 juin 2009, p. 2.

⁴³¹ John GERRING, *op. cit.*; Robert K YIN, *op. cit.*

Si les travaux de sociologie nous éclairent sur la pertinence du recours à l'étude de cas, c'est dans les travaux d'évaluation de politique publique que nous avons glané les jalons opérationnels de l'emploi de cette méthode. Marginale au début des années 80, elle a été peu à peu reconnue comme pratique légitime en évaluation, comme en témoigne un rapport méthodologique du General Accounting Office (équivalent américain de la Cour des Comptes), en 1990, dédié entièrement à l'étude de cas⁴³². Cette démarche requiert une construction méthodologique comprenant la définition des questions de recherche, le choix d'une unité de cas, d'un référentiel – c'est-à-dire d'un ensemble de questionnement et de points d'attention - et le choix des cas d'étude.

Les études de cas sont ici employées pour répondre à l'intégralité des questions de recherche et explorer les hypothèses présentées ci-dessus. L'unité des études de cas est la conduite d'une expérimentation par assignation aléatoire. Néanmoins cette recherche s'intéresse autant à l'amont de l'expérimentation - les raisons de l'investissement des différents acteurs dans cette expérimentation - qu'à l'aval – la valorisation et l'utilisation des résultats de cette dernière. La grille de questionnements utilisée pour aborder chaque cas était similaire, cependant l'accès aux informations ayant été difficile dans certains cas, nous avons approfondi chaque cas de manière variable.

Notre référentiel était le suivant :

Amont • Identification des parties prenantes de l'évaluation • Identification des motivations des parties prenantes pour mener une expérimentation et qu'elle soit évaluée par la méthode expérimentale • Modalités de rencontre des parties prenantes • Apprentissage souhaité par les parties prenantes • Négociations du design de l'expérimentation, sur le projet et les indicateurs pour estimer l'impact ainsi que le tirage aléatoire • Difficultés anticipées en amont de la mise en œuvre au vu du design Mise en œuvre • Difficultés rencontrées dans la mise en œuvre par les parties prenantes • Degré de contrôle des chercheurs sur la mise en œuvre • Degré d'éloignement de la mise en œuvre au regard du design établi • Apprentissages lors de la mise en œuvre Production des résultats • Modalités de production des résultats par les chercheurs • Degré de scientificité des résultats selon les chercheurs • Degré de pertinence des résultats pour les porteurs de projets Utilisation des résultats • Utilisation des résultats par les chercheurs • Utilisation des résultats par les porteurs de projets • Utilisation des résultats par les financeurs

Notre stratégie de sélection des cas a été organisée autour de plusieurs critères. D'abord les expérimentations devaient être évaluées par la méthode expérimentale par assignation aléatoire, sur le territoire français, et être finalisées au plus tard en 2012 afin que l'on puisse observer l'ensemble du processus et notamment l'utilisation des résultats. Du fait de la contemporanéité de notre objet de recherche, ce corpus qui se composait au début de notre recherche de moins d'une dizaine d'évaluation s'est peu à peu élargi au fur et à mesure des dépôts de rapports finaux. Aussi, beaucoup

⁴³² PROGRAM EVALUATION AND METHODOLOGY DIVISION, *Case Study Evaluations*, United States General Accounting Office, 1990; Robert K YIN, *op. cit.*

d'expérimentations ont changé de méthodes en cours de mise en œuvre voire ont été abandonnées faute de constituer un terrain propice à la production de résultats scientifiques. Nous étudierons par comparaison informelle quelques-uns de ces cas afin de comprendre les éléments sources de rupture des expérimentations qui nous permettront d'identifier en creux les facteurs relatifs au maintien d'autres expérimentations. Au fur et à mesure de l'élargissement de notre corpus, nous avons renoncé à un traitement exhaustif de l'intégralité des cas d'expérimentation aléatoire et nous avons opté pour une sélection de cas, basée sur la diversité des chercheurs et des porteurs de projet impliqués, afin de renforcer la validité externe de nos résultats. Nous avons sélectionné les cas d'expérimentation aléatoires initiées à partir des premières initiatives jusqu'au départ du M. Hirsch du gouvernement au printemps 2010. Toute expérimentation initiée après ne nous permettait pas d'étudier la prise en compte des résultats. Cette étude compte quatre cas d'expérimentations financées en dehors des initiatives du Haut Commissariat et 11 financées par les Hauts Commissariats (aux solidarités actives et à la jeunesse). Nous présenterons ces cas en détail dans le chapitre 3.

2.2.2.2. Contrôler la généralisation et la subjectivité de l'étude de cas

On reproche principalement à l'étude de cas son manque de rigueur, sa subjectivité ainsi que son incapacité à générer des enseignements généraux⁴³³. Nous proposons une stratégie de renforcement de la validité interne et externe de notre étude par une comparaison de plusieurs études de cas entre elles⁴³⁴. La généralisation est alors de nature analytique et non statistique. Contrairement à une démarche quantitative, les cas ne constituent pas des échantillons⁴³⁵. L'étude de cas est associée à ce que l'on appelle en méthodologie « la comparaison d'un nombre de cas restreints » (*small-N comparison*). En effet, l'étude de cas ne construit pas la validité de ses résultats sur la récurrence statistique de corrélations entre plusieurs variables. Une étude de cas est représentative au regard d'une théorie initiale qui la présente comme un point d'observation sélectionné pour un objet d'étude.

Cette opération requiert d'avoir recours aux outils de l'analyse comparée portant sur un petit nombre de cas, donc relevant davantage d'une démarche qualitative. En effet, pour tirer des conclusions, on ne peut utiliser, ni la méthode de concordance⁴³⁶, ni la méthode de différence⁴³⁷, proposée par John S. Mill, puisque le nombre limité de cas pourrait entraîner des interprétations erronées⁴³⁸. En d'autres termes, on ne peut conclure qu'une variable entraîne systématiquement la présence de l'effet voulu en ayant comparé la présence et l'absence de différentes variables dans une dizaine de cas.

⁴³³ Cette critique fut notamment formulée lors du conflit opposant les universitaires de Columbia et l'école de Chicago lors de la réunion annuelle de l'American sociological society en 1935, Jacques HAMEL, Stéphane DUFOUR et Dominique FORTIN, *Case study methods.*, London, SAGE, 1993, p. 23.

⁴³⁴ « the traditional small-scale community research of fieldwork anthropology are not in themselves generalizing studies. But they can easily become such if carried out in some numbers, so that judgments of their typicality can justifiably be made ». Anthony GIDDENS, *The constitution of society*, Berkley, University of California Press, 1984, p. 328.

⁴³⁵ Jacques HAMEL, Stéphane DUFOUR et Dominique FORTIN, *op. cit.* p. 44.

⁴³⁶ La méthode de concordance décrite par J.S. Mill est utilisée lorsque les résultats (variables dépendantes) des unités comparées sont semblables (par exemple personne en emploi). Elle suit le principe que si parmi toutes les unités, une variable est commune à toutes les unités, elle en est sûrement la cause John Stuart MILL, *Système de logique déductive et inductive.*

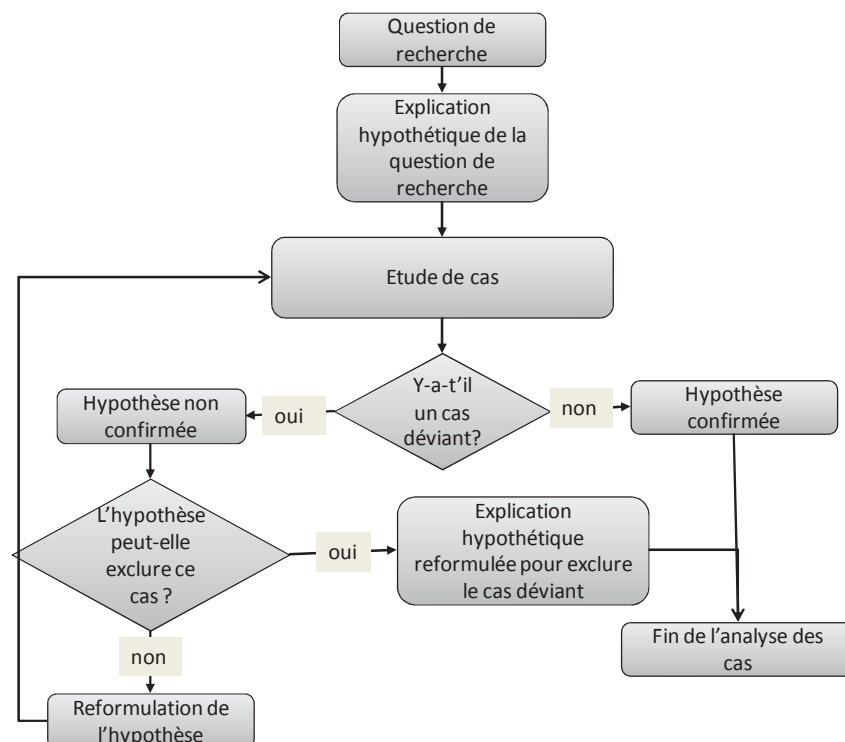
⁴³⁷ La méthode de différence décrite par J.S. Mill est utilisée lorsque les résultats (variables dépendantes) des unités comparées sont différents (par exemple personne en emploi, personne sans emploi). Elle suit le principe que si parmi toutes les unités une variable varie en parallèle avec le résultat final, elle en est sûrement la cause John Stuart MILL *Ibid.*

⁴³⁸ Stanley LIEBERSON, « Small N's and big conclusions: an examination of the reasoning in comparative studies based on a small number of cases », dans Charles C RAGIN et Howard Saul BECKER, *What is a case? : exploring the foundations of social inquiry*, Cambridge [England]; New York, NY, USA, Cambridge University Press, 1992, p. 105- 118.

Une solution alternative est possible par la structuration de la comparaison des cas autour de l'élaboration d'un modèle proposant une décomposition analytique des étapes de raisonnement constituant la théorie. Cette démarche consiste à faire émerger un schéma explicatif de la comparaison des études de cas entre elles et de modifier ce dernier jusqu'à ce qu'il puisse expliquer l'intégralité des cas. Cette démarche appelée induction analytique estime que le modèle est jugé comme ayant une validité externe⁴³⁹ lorsque l'on arrive à saturation, c'est-à-dire qu'on ne trouve plus de cas ne pouvant être expliqué par ce modèle⁴⁴⁰.

La méthode d'analyse par induction analytique a été théorisée par F. Znaniecki en 1935. En s'inspirant des méthodes utilisées en biologie et physique, cet auteur propose une méthode applicable aux sciences sociales et orientée vers la généralisation et la construction de théories à visées universelles. Cette méthode porte son attention sur les liens causaux amenant à un phénomène ou à une situation. Elle suit un processus itératif de confrontation des hypothèses de départ à une variété de cas et à une reformulation de ces dernières jusqu'à ce qu'elles puissent rendre compte de l'intégralité des cas⁴⁴¹. L'originalité de cette méthode consiste à fonder sa validité externe non pas sur la pertinence de son échantillon ou sur sa puissance statistique, mais sur l'atteinte d'une saturation des apports théoriques issus des cas. On ne cherche pas à exclure les cas marginaux mais au contraire à les utiliser comme ressources précieuses pour élaborer une théorie générale. Il n'y a donc pas un nombre nécessaire de cas prédéfinis, mais on décide d'arrêter quand les nouveaux cas ne viennent plus contredire en enrichir la théorie développée.

Figure 6: Logique de l'induction analytique



⁴³⁹ La validité externe est la capacité d'une démarche de recherche à fournir des résultats généralisables à d'autres populations, à d'autres contextes et à d'autres périodes

⁴⁴⁰ Howard BECKER, *Les ficelles du métier*, op. cit.

⁴⁴¹ Howard BECKER, *Les ficelles du métier*, op. cit. p. 323.

2.2.3. Protocole de collecte de données

Pour collecter les données, deux outils de collecte ont été mobilisés. Une analyse documentaire de l'ensemble des documents rendus public a été conduite en préalable (note d'étape, rapport d'évaluation, documents de presse, articles scientifiques). Elle a été alors complétée par une cinquantaine d'entretiens auprès de parties prenantes des expérimentations (évaluateurs, porteurs de projets et financeurs) en ayant pour souci de témoigner de la pluralité des positionnements institutionnels. Ces entretiens ont été aussi l'occasion de compléter l'analyse documentaire en sollicitant l'accès à des documents internes (compte rendu du comité de pilotage de l'évaluation, notes internes). Nous ne reviendrons pas ici sur la justification de l'emploi de l'entretien semi-directif à laquelle nous avons procédé pour le chapitre 2.

Cette collecte de données s'est déroulée entre février 2012 et novembre 2013. Malgré notre tentative de contacter l'intégralité des parties prenantes de tous les projets objets de nos cas d'étude, la rencontre des porteurs de projets s'est avérée très difficile. De nombreux porteurs n'ont pas souhaité se rendre disponibles et il a été parfois impossible d'identifier les personnes en charge du suivi du projet dans les structures porteuses. Nous avons tenté de compenser cet écueil par l'étude détaillée du maximum de documents relatifs au projet produits par les organisations ayant porté des expérimentations évaluées par la méthode expérimentale par assignation aléatoire. Par contre, la quasi-totalité des chercheurs/évaluateurs que nous avons souhaité rencontrer s'est rendue disponible pour répondre à nos interrogations et nous les en remercions. Nous précisons lors de chaque citation la position professionnelle de nos interlocuteurs ainsi que le projet auquel ils ont participé. Néanmoins, par souci de garder la confidentialité des propos, nous utilisons le masculin pour nous référer à chacun et chacune d'entre eux. Le faible nombre de femmes ayant participé aux expérimentations les rendrait trop facilement identifiables.

Ainsi, ce chapitre introductif, nous a permis d'éclairer notre stratégie de recherche abductive articulant l'observation du phénomène et l'élaboration du cadre théorique. L'expérimentation aléatoire apparaît comme un objet dont le périmètre d'utilisation est régulièrement redéfini selon l'interaction entre les acteurs la promouvant et la conjoncture particulière institutionnelle et contextuelle dans laquelle ils agissent. Son utilisation témoigne aussi de tensions au sein de cet instrument organisant l'interaction d'acteurs scientifiques et politiques, amenant à questionner son aptitude à organiser un apprentissage commun entre ces deux parties. Ces premiers constats nous ont amenée à élaborer un cadre théorique et une méthodologie nous permettant d'étudier ces deux aspects dans le contexte français. Nous proposons maintenant, de procéder à l'opérationnalisation de cet appareillage tout en mettant en perspective nos résultats empiriques avec la littérature présentée dans ce chapitre, afin d'en dégager les particularités mais aussi les points communs.

2^{ème} Partie : l'institutionnalisation de la «méthode expérimentale » en France: un succès temporaire ?

«La science politique se formera ; l'intrusion tardive mais certaine de la méthode scientifique dans l'art politique périmerà peu à peu les formes traditionnelles de Gouvernement, qu'elles soient démocratiques ou dictatoriales, et leur substituera des organes de direction à caractère technique, basés sur une connaissance indiscutable et claire de l'intérêt collectif, de l'intérêt mondial. [...] Le moment approche où la science économique portera ses premiers fruits et où, fortifié par cette nourriture indispensable à son développement, l'art politique deviendra une méthode rationnelle»

J. Fourastié, *La civilisation de 1960*, Paris, Presses universitaires de France, 1947 p.88

« Il s'agit d'appliquer à la politique une rigueur scientifique »

Esther DUFLO, *Introduction, Conférence nationale de l'expérimentation sociale, réformes contre placebo?*, Paris, Conférence nationale de l'expérimentation sociale, réformes contre placebo?, 29 mars 2010.

La première partie du chapitre précédent, s'est achevée chronologiquement à la fin des années 2000. Ce chapitre s'intéresse à la diffusion en France de la méthode expérimentale par assignation aléatoire comme instrument *d'evidence-based policy* et procède à un retour en arrière dans cette chronologie. Ces deux chapitres doivent donc davantage être analysés en parallèle, le chapitre premier pouvant apporter des éléments explicatifs de tendances observées dans le cas français. La problématique générale de ce travail s'intéressant aux modalités de reconfigurations de l'interaction entre l'action publique et les sciences sociales provoquée par la diffusion de la méthode expérimentale, il est essentiel dans ce chapitre de retracer aussi les modalités de cette interaction précédant l'emploi de cette méthode. De plus, nous avons vu dans le chapitre précédent, que les modèles *d'evidence-based policy* résultaient d'un équilibre entre un modèle standardisé et son interprétation par les acteurs scientifiques et politiques locaux, elle-même ancrée dans l'histoire nationale de leur interaction. Un retour historique est donc indispensable à ce chapitre.

Celui-ci s'intéresse au processus d'importation et de naturalisation de cette méthode d'évaluation sur le territoire français. Ce questionnement descriptif s'articule ici, à un questionnement de nature explicative : comment expliquer que cette méthode ait été si rapidement institutionnalisée comme étalon-or des méthodes d'évaluation par des acteurs de l'administration centrale française ? Alors que cette même administration avait été jusqu'à lors réfractaire à l'évaluation des politiques publiques conduite par des acteurs externes.

Nous avons dans le chapitre précédent présenté en détail le cadre théorique et méthodologique utilisé dans cette recherche. Notre perspective constructiviste de l'institution nous amène à étudier d'une part, les activités des acteurs dans le processus et d'autre part, à rendre compte des différents cadres et moteurs orientant leurs actions. Elle nous amène aussi à expliquer pourquoi certains acteurs deviennent pendant une conjoncture particulière, plus compétents que d'autres, pour agir sur ces cadres, les modifier et former une nouvelle institution. Pour ce faire, nous mobilisons les concepts de *champ d'action stratégique* et de *travail institutionnel*. Le premier, ouvre un programme de recherche sur les motivations des acteurs selon leurs champs d'appartenances et leur situation au sein de ces derniers. Le second, nous permet d'étudier les actions de création institutionnelle des acteurs en nous intéressant aux acteurs clés, les *entrepreneurs*, mais aussi aux autres acteurs influençant les modalités de l'incarnation de cette institution.

Nous avons développé les quatre hypothèses suivantes :

- *Hypothèse. 2.1 : L'évaluation expérimentale par assignation aléatoire a constitué une ressource pour faire face à des crises autant dans le champ de l'action publique que de la science économique appliquée en accordant de nouveaux avantages de positionnement à ses promoteurs.*
- *Hypothèse. 2.2 : Cette méthode doit son succès en France, à la mobilisation des acteurs traditionnels de l'expertise de l'administration centrale et s'inscrit en partie dans une logique prospective et non rétrospective correspondant à la logique française de mobilisation des savoirs.*

Une fois que nous aurons analysé, d'une part, les structures sous-jacentes à l'interaction des sciences sociales et de l'action publique au sein de l'Administration centrale française (*Hyp. 2.2*), ainsi que les motivations pouvant expliquer l'organisation d'un travail institutionnel (*Hyp. 2.1*) autour de la constitution d'un *champ d'action*

stratégique promouvant cette méthode, nous aborderons alors le travail d'agence des promoteurs de cette méthode et du courant de *l'evidence-based policy* en France autour d'une dernière hypothèse :

- *Hypothèse 2.3 : La promotion de cette méthode a bénéficié d'un travail institutionnel intense conduit par un champ d'action stratégique temporaire alliant acteurs scientifiques et acteurs de l'action publique.*

Ces éléments précisés, nous pouvons procéder maintenant à la présentation des résultats de cette recherche. Ces derniers sont présentés suivant un plan chrono-thématique permettant au lecteur de suivre, tel un récit, l'histoire du développement de cette méthode en France, mais aussi d'identifier, à travers lui, les éléments pouvant être éclairés par l'échafaudage théorique élaboré ci-avant. Nous situant dans une épistémologie réaliste, le résultat n'est pas d'aboutir à une validation ou invalidation de cette théorie, mais ici de mettre en lumière son apport pour identifier les structures sous-jacentes et les mécanismes à l'œuvre dans ce processus, tout en accordant une place déterminante au contexte national auquel il s'applique.

D'abord, nous étudions les racines de l'émergence du *champ d'action stratégique*. Cette partie souligne l'inscription du travail institutionnel dans la tradition de l'organisation de la rencontre des sciences sociales et de l'État en France, à l'écart du développement de l'évaluation des politiques publiques. Elle revient ensuite, sur les actions d'agence du champ stratégique embryonnaire, et notamment les premières initiatives d'organisation ainsi que la définition du discours commun.

Puis, nous présentons la stratégie de recherche d'alliés notamment lors de l'accession aux instances de pouvoirs d'une partie du *champ d'action stratégique* entre 2007 et 2009. Nous nous intéressons d'abord, à la stratégie du pendant politique du champ d'action stratégique en revenant sur le travail du cabinet de Martin Hirsch, plus particulièrement sur l'expérience du Revenu de Solidarité Active, puis le premier appel à projets à expérimentations. Ensuite, nous abordons le travail effectué par le pendant scientifique du champ, à travers la mobilisation des alliés dans le domaine de l'emploi ainsi que dans les sphères scientifiques.

Enfin, nous revenons sur la stratégie d'institutionnalisation de ce *champ d'action stratégique*, à travers la création du Fonds d'Expérimentations pour la Jeunesse et, en son sein, d'une *unité de gouvernance interne* ayant pour objectif de stabiliser le champ. Il apparaît que cette stratégie a été victime de la déstabilisation du champ par la disparition de son pendant politique suite à la dissolution du cabinet de Martin Hirsch ainsi que du manque de *suiveurs* s'inscrivant dans la définition du champ stratégique par les *titulaires*. Cette institution a été ensuite la proie des *champs d'action stratégiques* gouvernementaux le détournant de sa raison d'être originelle.

En conclusion, il apparaît que, malgré son développement modeste, cette méthode s'est accompagnée, pour un temps, d'une reconfiguration fondamentale des relations entre des acteurs scientifiques et ceux de l'action publique au sein d'un champ d'action stratégique temporaire et restreint, puis à été peu à peu dissoute par le travail institutionnel de *challengeurs* et la perte des ressources politiques de ses *titulaires*.

1. La conjoncture française de l'émergence d'un champ d'action stratégique de la méthode expérimentale

Ce premier développement empirique est consacré à l'étude des éléments du contexte historique français composant un terreau favorable, voire pouvant composer la source des motivations de la constitution d'un *champ d'action stratégique* autour de la promotion de la pratique de la méthode expérimentale par assignation aléatoire. Si, de prime abord, au vu de l'histoire de l'évaluation en France, la promotion de cette méthode et son adoption par des instances de l'État peuvent paraître paradoxales, sa mise en perspective avec l'histoire de l'organisation de la rencontre de la science économique et de l'action publique révèle son inscription dans la tradition française d'organisation de l'expertise d'État. L'intérêt pour cette méthode n'est pas nouveau en France, et celle-ci a d'ailleurs constitué un fil rouge dans l'histoire des courants nationaux de l'évaluation. Ce terreau propice à cette méthode est alors « activé » dans les années 2000 par un groupe de scientifiques ayant identifié cette dernière comme une ressource permettant un nouveau positionnement dans le champ scientifique français et international. Ce travail institutionnel prend une dimension organisationnelle nouvelle par la rencontre avec des acteurs politiques et administratifs l'identifiant aussi comme une ressource dans leur propre champ. C'est alors que se dessinent les contours d'un nouveau *champ d'action stratégique*, ainsi que la production d'un discours de problématisation commun présentant la méthode expérimentale comme une innovation répondant aux problèmes rencontrés par ces champs respectifs.

Nous présentons dans un premier temps, un retour sur la structuration des relations entre l'action publique et les acteurs proposant des méthodes scientifiques pour l'éclairer cette approche en France. Cette association au sein de l'administration centrale est caractérisée, depuis la fin de la seconde guerre mondiale, par un monopole de cette expertise exercée par un cercle d'acteurs privilégiés, d'économistes statisticiens appliqués ou de Grands Corps comme notamment la Cour des comptes. Ce monopole a constitué notamment, un des obstacles à l'institutionnalisation de l'évaluation des politiques publiques au sein de l'administration centrale. L'expérimentation quant à elle, n'est pas un objet apparu dans les années 2000 en France. Néanmoins, jusqu'à présent, ses pendants scientifiques et politiques ont été développés séparément. On observe d'une part, plusieurs initiatives politiques innovantes mais sans pour autant qu'elles soient associées à une méthode d'évaluation particulière et d'autre part, un intérêt grandissant des économistes statisticiens appliqués pour cette méthode de démonstration scientifique. Après avoir établi ce contexte, nous revenons sur la conjoncture particulière du début de la décennie 2000 ayant permis l'émergence d'un champ d'action stratégique embryonnaire composé d'acteurs politiques et scientifiques promouvant le recours à cette méthode d'évaluation comme instrument de *evidence-based policy*.

1.1. Inscription du *travail institutionnel* dans l'histoire de l'organisation de la rencontre des méthodes des sciences économiques et de l'Administration centrale en France, à l'écart du champ de l'évaluation des politiques publiques

La rapidité avec laquelle la haute administration ainsi que certaines instances de l'État se sont saisies de la méthode expérimentale, ne peut être attribué uniquement à l'impulsion donnée par le travail institutionnel de quelques individus. Ce premier succès s'explique en partie par l'inscription de cette méthode, par sa logique et le type d'expertise qu'elle mobilise, dans la tradition française de rencontre entre la science économique et l'action de l'État. *A contrario*, l'histoire de l'évaluation des politiques publiques en France, qui s'est écrite principalement à l'écart de l'Administration centrale, s'est constituée en opposition à la méthode expérimentale. Bien que « l'expérimentation » soit une pratique répandue en France, notamment dans une logique de décentralisation, celle-ci ne s'accompagnait pas d'un pendant évaluatif particulier. Cette méthode apparaît néanmoins en filigrane en France comme structurante, de manière dormante, dans la boîte à outils méthodologiques des « économistes appliqués »⁴⁴².

Ce détour historique permet de rendre compte de la structuration des champs dont sont issus les entrepreneurs de l'expérimentation aléatoire, champs qui subiront ensuite, l'influence du travail institutionnel engagé par ces derniers. Nous observons d'une part, l'existence d'un champ d'action stratégique de l'expertise d'Etat dont les titulaires sont les économistes appliqués de l'INSEE (Institut national des statistiques et des études économiques) et d'autre part, un champ de l'évaluation, déserté par les chercheurs, à l'écart de l'administration centrale et dont les titulaires sont des cabinets privés de conseil. Si l'on repère, dès la fin des années 90, un intérêt de quelques acteurs du champ d'expertise d'Etat pour la méthode expérimentale par assignation aléatoire qui apparaît comme une ressource dans le champ stratégique international de l'économie appliquée, celui de l'évaluation des politiques publiques s'est, quant à lui, développé dans les années 90 en opposition à cette méthode, en proposant une évaluation davantage ancrée dans la sociologie de l'action publique. L'expérimentation, au sens d'initiative d'action publique temporaire sur une partie du territoire est déjà présente sur le territoire français mais son usage reste marginal, et n'est pas encore lié à une méthode d'évaluation particulière.

1.1.1. L'association entre les sciences sociales et l'action publique en France : l'expertise à l'écart des sentiers de l'évaluation

Ce développement n'a pas pour visée de rendre compte de manière exhaustive de l'histoire de l'interaction entre l'action publique et les sciences sociales en France. Ce travail a déjà été conduit par ailleurs et cette étude n'apporte que très peu de nouvelles

⁴⁴² Ce sont ces économistes qui se définissent comme tels, ils entendent par là « qu'ils ont une démarche empirique d'analyse de données statistiques sur des situations particulières »

données primaires à ce sujet⁴⁴³. Le travail ci-dessous propose d'extraire de ces études les éléments indispensables à la compréhension des enjeux et de l'héritage à l'œuvre dans l'émergence de la méthode expérimentale par assignation aléatoire en France. Il semble indispensable de commencer par un détour historique revenant sur les modalités d'organisation de la rencontre entre sciences humaines et action publique en France en questionnant son lien avec l'évaluation des politiques publiques afin de mieux comprendre, par la suite, les motivations des acteurs, ainsi que les changements engendrés par la promotion et la conduite de la méthode expérimentale par assignation aléatoire. La présentation des choix de modalités de recours à l'expertise, par l'administration centrale française, est un élément essentiel à la compréhension de la faible institutionnalisation de l'évaluation des politiques publiques en France, ainsi que de la faible participation des scientifiques français à cette activité.

1.1.1.1. L'organisation de l'expertise d'État, ou le monopole de la prospective économique et des Grands Corps, comme obstacle à l'évaluation des politiques publiques

C'est dans l'après-guerre que commence à s'organiser le recours à l'expertise scientifique par l'État français sur le modèle américain⁴⁴⁴. L'État, qui doit organiser la restructuration de l'action publique et faire face à une croissance exceptionnelle de la population, doit répondre à un besoin d'expertise sans précédent. Le Commissariat général au Plan, créé en 1945, est le lieu privilégié de cette rencontre entre science économique et action d'État. Certains ont pu identifier, à partir de son évolution au début des années 1950, la planification comme un mode de gouvernement rationnel de l'action publique⁴⁴⁵. Les modalités d'interaction entre sciences sociales et action de l'État posées par cette instance ont été structurantes jusqu'à aujourd'hui. V. Spenlehauer revient sur cette entreprise, en expliquant qu'une « *communauté de planificateurs* » se met en place dans les années 1960 entre le personnel du Plan et l'INSEE, nouvellement créée⁴⁴⁶. L'INSEE a notamment en charge, de générer des modélisations macro-économiques permettant d'estimer l'impact des orientations envisagées par le Commissariat général. Ce commissariat organise aussi l'interface avec le milieu de la recherche, en témoigne notamment la création du Comité d'organisation des recherches appliquées sur le développement économique et social (CORDES). Des comités d'experts réunissent des polytechniciens, des chercheurs, dont des sociologues comme M. Crozier et R. Aron, afin de mieux comprendre les enjeux du changement de la croissance économique sur la société française⁴⁴⁷. Ces sociologues, dont Paul-Henri Chombart de Lauwe, questionnent la légitimité de l'intervention du scientifique dans les affaires d'État, soulignant la différence notoire entre la société « *telle qu'elle est* » et

⁴⁴³ Voir notamment la thèse de Vincent Spenlehauer, « L'évaluation des politiques publiques, avatar de la planification » (Université Pierre Mendès-France, 1998) ou l'ouvrage de Rouban (L.), 1988, L'État et la science : la politique publique de la science et de la technologie, Paris, ou Jobert (B.), 1981, Le social en plan, Paris, Les Editions Ouvrières Editions du CNRS, ou même BEZES P., et al., L'État à l'épreuve des sciences sociales. La fonction recherche dans les administrations sous la Ve république, Paris, La Découverte, 2005.

⁴⁴⁴ Jean Monnet promeut la création du commissariat général au plan « avec l'appui technique et stratégique d'experts et universitaires mandatés par le *Council of Economic Advisers* de la présidence américaine et par le *Policy planning staff* du département d'État américain. Vincent SPENLEHAUER, « Intelligence gouvernementale et sciences sociales », (1999) 12-48 *Politix* 95-128, p. 96.

⁴⁴⁵ *Ibid.* p.103.

⁴⁴⁶ Les statisticiens de l'INSE sous l'impulsion de C. Gruson (ancien membre du Service des études économiques et financières qui devient directeur de l'INSEE en 1961) sont orientés vers des activités de comptabilité nationale et de planification économique une culture du « bidouillage » qui peut rentrer en conflit avec la préoccupation de la qualité technique des enquêtes statistiques, courant prédominant à l'INSEE qui déterminera aussi en partie l'intérêt pour la méthode expérimentale. Pour un historique détaillée de la relation du commissariat au plan et de l'INSEE se référer au travail de Vincent SPENLEHAUER, *L'évaluation des politiques publiques, avatar de la planification*, op. cit. p. 34.

⁴⁴⁷ Vincent SPENLEHAUER, *L'évaluation des politiques publiques, avatar de la planification*, op. cit. p. 43.

celle « *que les planificateurs voudraient qu'elle soit* ». La position difficile du sociologue vis-à-vis du pouvoir est aussi soulignée en précisant que ce dernier doit attirer l'œil du pouvoir sur un problème de société et non pas accepter des sujets de recherche imposés par le pouvoir. Ce sociologue souligne qu'« *il faut instaurer un dialogue, et non simplement passer commande* »⁴⁴⁸.

À partir de cette période, s'installe un monopole de la prospective économique sur le conseil de l'Administration centrale, et une distanciation d'une majorité de sociologues qui refusent de répondre aux commandes d'État. Les membres du Commissariat sont teints d'une technocratie manifeste où des scientifiques, exclusivement des économistes, guident l'action publique. V. Spenlehauer souligne que ces acteurs restent ancrés dans la prospective et n'ont pas d'intérêt pour l'évaluation *ex post* de l'effectivité des orientations économiques qu'ils préconisent. S'amorce aussi un dénigrement de la politique partisane en faveur de la technique. Delphine Dulong souligne qu'au cours de la Ve république on observe une intervention croissante des économistes en tant que conseillers de l'action publique, lesquels se présentent en technocrates, figure en « *néгатif* » - au sens photographique - de la figure discréditée du parlementaire. La compétence économique, selon cette auteure, est peu à peu convertie en ressource politique⁴⁴⁹. L'économiste est présenté comme ne tenant pas compte des enjeux électoraux et n'étant pas aux prises de l'idéologie⁴⁵⁰. Cette figure n'est pas sans rappeler celle défendue par les acteurs se saisissant de la méthode expérimentale par assignation aléatoire.

En 1967, naît le mouvement de la Rationalisation des Choix Budgétaires, importation du modèle américain du *PPBS* (*Planning, Programming, and Budgeting System*) qui se présente comme la « *recherche scientifique et extensive de toutes les procédures et de toutes les méthodes qui permettraient d'améliorer la pertinence, la cohérence et l'efficacité des décisions publiques pourvues de moyens budgétaires, ou dont l'effet aurait sa trace dans le Budget* »⁴⁵¹. Ce mouvement s'inscrit déjà dans un type d'*evidence-based policy*, au sens où il prétend restaurer la légitimité de l'administration par l'application de méthodes scientifiques. Cette expérience est abandonnée en 1985, mais elle contribue à mettre en place une systématisation de la collaboration entre les acteurs décisionnels et des acteurs du calcul de prospective économique⁴⁵². L'échec de la RCB, et son démantèlement à partir de 1974, ouvrent la voie en 1980, à l'évaluation des politiques publiques⁴⁵³. La France accuse un retard notoire dans ce domaine par rapport aux pays anglo-saxons. J.-P. Nioche et R. Poinsard, lors du colloque qu'ils ont organisé en 1983 à l'ENA (École Nationale d'Administration), recensent en France uniquement 42 travaux d'évaluations *ex post* en 1983, et moins de 10 évaluations par an jusqu'en 1976⁴⁵⁴.

⁴⁴⁸ P.-H. Chombart de Lauwe cité par Spenlehauer Vincent SPENLEHAUER, *L'évaluation des politiques publiques, avatar de la planification*, Grenoble, Université Pierre Mendès-France, 1998, p. 44.

⁴⁴⁹ Delphine DULONG, « Quand l'économie devient politique. La conversion de la compétence économique en compétence politique sous la Ve République », (1996) 9-35 *Politix* 109-130, p.118.

⁴⁵⁰ On peut faire l'hypothèse que ce mouvement français vers un État technocrate s'effectue en réaction à une 4^{ème} République aux prises des alliances partisans d'un régime fortement parlementaire source d'instabilité gouvernementale et ayant finalement accordé les peins pouvoirs au Maréchal Pétain.

⁴⁵¹ Romain LAUFER, *Le prince bureaucrate: Machiavel au pays du marketing*, Paris, Flammarion, 1982, p. 141.

⁴⁵² Jean-Claude BARBIER et Nicolas MATYJASIK, « Évaluation des politiques publiques et quantification en France : des relations ambiguës et contradictoires entre disciplines », (2010) 1-5 *Revue Française de Socio-Economie* 123-140, p.133.

⁴⁵³ Ce lien est explicité par l'organisation du colloque « où va la gestion publique ? » à HEC avec pour but d'envisager un après RCB, au cours duquel J-P Nioche propose de construire un nouveau mode d'expertise autour l'évaluation des politiques publiques Vincent SPENLEHAUER, « Intelligence gouvernementale et sciences sociales », *op. cit.* p. 121.

⁴⁵⁴ Frank LEVY, « L'évaluation des politiques publiques aux Etats-Unis », dans Jean-Pierre NIOCHE et Robert POINSARD, *L'Evaluation des politiques publiques*, Paris, Economica, 1984, p. 27-33 p. 8.

De premières initiatives évaluatives étatiques se dessinent avec la rédaction en 1986, d'un rapport dit « Deleau », rédigé par le Commissariat au plan. L'évaluation d'État gagne en visibilité avec l'évaluation du RMI (Revenu minimum d'insertion), première mesure expérimentale de grande ampleur de 1988 à 1992⁴⁵⁵. Mais l'impulsion notoire survient dans les années 1990 avec la circulaire Rocard du 23 février 1989, laquelle va susciter l'émergence d'un ensemble disparate d'initiatives en faveur de l'évaluation des politiques publiques, dont certaines menées par l'État⁴⁵⁶. Le décret du 22 janvier 1990 définit l'évaluation en ces termes : « *évaluer les politiques publiques consiste à mesurer les effets qu'elles engendrent et à chercher si les moyens juridiques, administratifs et financiers mis en œuvre produisent les effets qu'on en attend* »⁴⁵⁷. Celle-ci est inscrite dans une double démarche, démocratique d'une part, d'information des citoyens et, d'autre part, de réforme de l'État, en exerçant « *une pression analogue à celle que le marché fait peser sur les entreprises* »⁴⁵⁸. Par le même décret sont créés le Comité Interministériel de l'Évaluation (CIME) qui décide des projets - soumis par les ministères - à évaluer ; un Fonds national de développement de l'évaluation ; et le Conseil scientifique de l'évaluation (CSE) qui fonctionne comme l'auxiliaire scientifique du CIME, responsable du contrôle méthodologique des évaluations sélectionnées. Le CSE, qui réunit en son sein des membres des grands corps d'inspection et de l'INSEE, de la Cour des comptes et des universitaires et experts, travaille à la production d'une doctrine française en évaluation⁴⁵⁹. Au-delà du travail de définition méthodologique, que l'on retrouve notamment dans le guide produit à cet effet, son activité de contrôle de travaux d'évaluation est très limitée et tombe peu à peu en désuétude puisque, de 1990 à 1996, seules une vingtaine d'évaluations ont été suivies⁴⁶⁰. Le Conseil scientifique fut remplacé, en 1998, par le Conseil national de l'évaluation (CNE) qui, à son tour, tombe en désuétude et ne sera jamais renouvelé⁴⁶¹.

Ces difficultés d'institutionnalisation s'expliquent par une pluralité d'éléments conjoncturels et structurels. D'abord, comme le souligne M. Rocard, elle a été victime d'un « *réflexe généralisé de méfiance* » envers une pratique se rapprochant, dans l'imaginaire des agents administratifs, du contrôle comptable, sans pour autant en être, et surtout, sans être menée par la Cour des comptes, mais par des instituts privés extérieurs à l'administration⁴⁶². En effet, Jean Leca, ancien président du Conseil Scientifique de l'Évaluation, estime que l'évaluation a été vécue par les grands corps d'État comme une menace à l'autorité de l'expertise officielle qui s'opérait dans la

⁴⁵⁵ Jean-Claude BARBIER et Nicolas MATYJASIK, *op. cit.* p. 125.

⁴⁵⁶ Jean-Claude BARBIER, « Éléments pour une sociologie de l'évaluation des politiques publiques en France », (2010) 1-2 *Revue Française des Affaires Sociales* 27-49, p.29.

⁴⁵⁷ Décret du 22 janvier 1990 relatif à l'évaluation des politiques publiques, (1990), article 1.

⁴⁵⁸ Ces deux dimensions sont expliquées en ces termes l'évaluation « *répond (...) à une exigence de la démocratie en ce qu'elle permet de donner aux citoyens des informations et des appréciations objectives qui leur permettent de mieux fonder leur jugement sur la marche des services publics* ». « *elle constitue aussi un élément de modernisation de l'État, car une meilleure connaissance de l'efficacité réelle de l'action de l'administration exerce sur cette dernière une pression analogue à celle que le marché fait peser sur les entreprises* ». Vincent SPENLEHAUER, « Intelligence gouvernementale et sciences sociales », *op. cit.* p.124.

⁴⁵⁹ Jean-Claude BARBIER, « Éléments pour une sociologie de l'évaluation des politiques publiques en France », *op. cit.* p.33.

⁴⁶⁰ CONSEIL SCIENTIFIQUE DE L'ÉVALUATION, *Petit guide de l'évaluation des politiques publiques*, Paris, La Documentation française, 1996. Le nombre limité d'évaluations ne vient pas du manque de projet d'évaluation soumis par les ministères, mais vient en partie par la régulation par le CIME des projets sélectionnés. V. Spenlehauer souligne que par exemple en 1990 P. Moscovici au secrétariat du CIME n'a sélectionné pour évaluation aucun des 35 projets d'évaluation proposés par le ministère de l'Éducation nationale et de la recherche. Vincent SPENLEHAUER, « Intelligence gouvernementale et sciences sociales », *op. cit.* p. 126.

⁴⁶¹ Jean-Claude BARBIER et Nicolas MATYJASIK, *op. cit.* p. 125.

⁴⁶² Michel ROCARD, *Les nouvelles frontières de l'évaluation*, Marseille, Journées de la Société Française d'Évaluation, 11 juin 2009.

dispense d'expliciter les points de vue et les méthodes permettant de porter un jugement⁴⁶³.

Cette difficile institutionnalisation tient aussi, en partie, à l'opposition entre la culture d'évaluation managériale et la culture d'évaluation démocratique, la première étant incarnée par le rapport dit « Deleau », la seconde par le rapport dit « Viveret »⁴⁶⁴. Le premier rapport, rédigé dans le cadre des travaux du Plan, met en avant deux règles fondamentales, « la reproductibilité » des protocoles de l'évaluation, et la « séparation » de l'évaluateur par rapport aux agents responsables de la politique évaluée, afin de garantir l'impartialité de l'évaluation⁴⁶⁵. Il s'inscrit dans une logique de sortie de l'État-providence et il a une visée d'optimisation des moyens budgétaires. Le rapport Viveret quant à lui s'inscrit dans des inquiétudes liées à un contexte de crise de dispositifs de représentation traditionnels⁴⁶⁶. Ce débat n'est d'ailleurs pas particulièrement français, puisqu'il correspond aux conflits entre la troisième et quatrième génération d'évaluations décrites dans le contexte anglo-saxon par E. Guba et Y. Lincoln⁴⁶⁷. Ces controverses sont structurantes dans le débat français et distinguent d'un côté ceux qui estiment, dans la tradition du Commissariat au plan, qu'évaluer correspond à estimer une valeur, et ceux pour qui l'évaluation est un processus de médiation constructeur de consensus entre les parties prenantes de la politique évaluée.

En parallèle à ces initiatives, des instances d'institutionnalisation internes de l'interaction entre les sciences sociales et l'action publique s'organisent au sein de l'administration centrale française. On note la création de la Mire⁴⁶⁸ (Mission Interministérielle Recherche-Expérimentation) qui constitue un exemple de collaboration de la recherche et de l'action publique en France. L'action de la Mire consiste alors à évaluer les besoins de recherche liés à la définition et à la mise en œuvre des politiques en matière de travail et d'emploi, d'action sociale, de protection sociale et de santé⁴⁶⁹. Elle organise déjà, la rencontre entre des laboratoires, des établissements publics de recherche et des acteurs politico administratifs. Néanmoins, son objectif n'est pas d'évaluer les mesures gouvernementales. L'évaluation s'y déroule « *au risque de la recherche* » et non l'inverse⁴⁷⁰. Comme le souligne V. Spenlehauer, la mission de ces structures est surtout de mener des travaux de recherche et non pas d'analyses, puisque ce sont les structures qui les commanditent qui ont ce rôle de traduction entre le savoir et son application⁴⁷¹. En 1998, la MIRE a intégré la DREES, (Direction de la recherche, des études, de l'évaluation et des statistiques) en devenant la Mission de la recherche.

⁴⁶³ Jean LECA, « L'évaluation dans la modernisation de l'État », (1993) 11-2 *Politiques et Management Public*.

⁴⁶⁴ Jean-Claude BARBIER, *op. cit.* p. 40.

⁴⁶⁵ Michel DELEAU, *Evaluer les politiques publiques: méthodes, déontologie, organisation: rapport du Groupe de travail « Méthodes d'évaluation des politiques publiques »*, coll. Plan / Commissariat général du plan, n°13, Paris, Documentation française, 1986.

⁴⁶⁶ Patrice DURAN et Eric MONNIER, « Le développement de l'évaluation en France. Nécessités techniques et exigences politiques », (1992) 42-2 *revue française de science politique* 235-262.

⁴⁶⁷ Dans leur travail ces auteurs identifient quatre générations de démarches évaluatives la première de nature technique, la seconde descriptive la troisième orientée vers le jugement, impliquant l'étude de l'atteinte des objectifs, de la mise en place de standards et pour laquelle l'évaluateur se positionne comme juge et la quatrième orientée vers les parties prenantes où l'évaluateur agit comme un médiateur dans le processus de négociation des intérêts. E GUBA et Y LINCOLN, « The countenances of 4 generation evaluation : Description, judgement and negociation », dans Dennis James PALUMBO, *The Politics of program evaluation*, *op. cit.*

⁴⁶⁸ Sur ce modèle de la MIRE au ministère des affaires sociales sont développées au début des années quatre-vingt le PUCA à l'Équipement, et l'IHESI à l'intérieur.

⁴⁶⁹ Jean-Noël CHOPART et Claude MARTIN, *Sciences de l'Homme et de la société et évaluation des politiques publiques, Evaluation et Décision, séminaire de méthode*, coll. Plan Urbain, CEOPS, Ecole Nationale des Travaux Publics de l'Etat, 19 septembre 1990, p. 139- 149, p. 149.

⁴⁷⁰ *Ibid.* p. 139.

⁴⁷¹ Vincent SPENLEHAUER, « Intelligence gouvernementale et sciences sociales », *op. cit.* p. 97.

Plusieurs Services Statistiques Ministériels (SSM), mélangeant les activités d'évaluation, de recherche et de prospectives sont créées comme la DEPP (Direction de l'évaluation, de la prospective et de la performance) au ministère de l'Éducation nationale à la fin des années 1980. À son image, mais dans le domaine de l'emploi est créé la DARES (Direction de l'animation de la recherche, des études et des statistiques) en février 1992, puis la DREES pour la Direction des affaires sociales et de la santé. Ces instances organisent l'interaction des services de l'Administration centrale avec les sciences sociales appliquées, et notamment les statistiques. Ces services sont composés principalement d'administrateurs de l'INSEE.

La collaboration des sciences sociales et de l'action publique de l'État est marquée historiquement par une logique de prospective et de statistique, à l'écart de l'évaluation. Quand évaluation il y a, au sein de ces instances, elle s'appuie sur des modèles économiques fondés sur des hypothèses de comportements des agents. L'évaluation tient alors d'une approche structurelle et prend la forme d'une simulation *ex ante* des changements de décision des agents résultant de cette réforme⁴⁷². Les questionnements *ex post* portent avant tout sur l'évaluation des réalisations plus que sur celle des résultats ou des impacts de l'intervention publique⁴⁷³.

Ainsi les observateurs de l'évaluation soulignent unanimement que la France accuse un retard dans le développement au sein de l'administration centrale de l'évaluation des politiques publiques⁴⁷⁴. La France a pour particularité nationale une centralisation de l'expertise qui révèle un quasi-monopole des statistiques, des études, et de l'expertise, par l'INSEE ainsi que par les Grands Corps⁴⁷⁵. L'administration centrale est considérée comme garante de « l'intérêt général » par nature et donc légitime de ce monopole du savoir⁴⁷⁶. Comme l'ont souligné P. Laborier et D. Gaxie, il semble que le potentiel critique de l'évaluation, ou la « zone d'autonomie » qu'elle requiert pour reprendre les termes de J. Leca -, ait été jusqu'à présent le principal obstacle à son développement. Selon J-C Barbier, « pour que l'évaluation se développe en tant que champ particulier d'une profession, concurrente et complémentaire à la fois du contrôle ou de l'inspection, sur la base de standards indépendants, il faudrait logiquement que ces acteurs traditionnels y trouvent un intérêt⁴⁷⁷ ». La méthode expérimentale par assignation aléatoire apparaît comme l'outil suscitant justement l'intérêt des acteurs traditionnels de l'expertise pour l'évaluation, ouvrant une diffusion rapide de cette dernière.

⁴⁷² François Bourguignon résume cette démarche dans son intervention lors d'un colloque de la DARES, *Colloque sur les expérimentations pour les politiques publiques de l'emploi et de la formation*, Paris, 22 mai 2008, p. 168, p. 145.

⁴⁷³ Daniel GAXIE et Pascale LABORIER, « Des obstacles à l'évaluation des actions publiques et quelques pistes pour tenter de les surmonter », dans *Etre gouverné: études en l'honneur de Jean Leca*, Paris, Presses de Sciences po, 2003, p. 201-224.

⁴⁷⁴ Philippe CORCUFF, « Un OVNI dans le paysage français ? Eléments de réflexion sur l'évaluation des politiques publiques en France », (1993) 6-24 *polix* 190- 209; Jean-Pierre NIOCHE, « De l'évaluation à l'analyse des politiques publiques », (1982) 32-1 *Revue Française de Science Politique* 32- 61; Daniel GAXIE et Pascale LABORIER, « Des obstacles à l'évaluation des actions publiques et quelques pistes pour tenter de les surmonter », dans *Etre gouverné: études en l'honneur de Jean Leca*, Paris, Presses de Sciences po, 2003, p. 201- 224; Jean-Claude BARBIER et Nicolas MATYJASIK, *op. cit.*; Frédéric VARONE et Steve JACOB, « Institutionnalisation de l'évaluation et Nouvelle Gestion Publique: un état des lieux comparatif », (2004) 11-2 *Revue Internationale de Politique Comparée* 271- 292; Vincent SPENLEHAUER, *L'évaluation des politiques publiques, avatar de la planification*, *op. cit.*; Patrice DURAN et Eric MONNIER, *op. cit.*; Clément LACOUETTE-FOUGERE et Pierre LASCOUMES, « Les scènes multiples de l'évaluation » *Les problèmes récurrents de son institutionnalisation*, Laboratoire interdisciplinaire d'évaluation des politiques publiques, Sciences-po Paris, 2003.

⁴⁷⁵ Cette particularité notamment vis-à-vis des États-Unis est notée dans le rapport du sénateur Bourdin mentionné par Jean-Claude BARBIER, *op. cit.* p. 37.

⁴⁷⁶ Jean-Claude BARBIER et Nicolas MATYJASIK, *op. cit.* p. 128; Jean-Pierre NIOCHE, « De l'évaluation à l'analyse des politiques publiques », (1982) 32-1 *Revue Française de Science Politique* 32-61, p.43.

⁴⁷⁷ Jean-Claude BARBIER, *op. cit.* p. 38.

1.1.1.2. Le développement du champ de l'évaluation à l'écart de l'Administration centrale et des milieux scientifiques

L'Administration centrale française ainsi que les milieux scientifiques français accusent un retard dans la pratique et l'institutionnalisation de l'évaluation des politiques publiques. Bien que cette pratique se soit développée dans d'autres sphères depuis les années 1990. Les collectivités territoriales comme les Conseils Régionaux, ou Généraux, ont été un terreau fertile au développement de la demande en évaluation. Quant à l'offre d'évaluation, elle s'est constituée à la marge des milieux scientifiques, au sein de petits cabinets de conseils dédiés à cette activité. Cette composition de l'offre d'évaluation a eu pour tendance d'orienter le débat évaluatif français, avant les années 2000, davantage sur la pertinence de la démarche d'évaluation vis-à-vis du contexte politique dans lequel elle s'inscrit, que sur les exigences de rigueur scientifique des méthodologies d'évaluation.

Le partenariat et la contractualisation constituent des facteurs majeurs du développement de l'évaluation des politiques publiques, notamment au sein des collectivités territoriales. Dans le cadre de l'obligation de l'évaluation des Contrats-Plan État-Région en 1993, plus de 150 évaluations ont été réalisées entre 1994 et 1999⁴⁷⁸. L'intérêt général étant représenté dans l'imaginaire politique français comme incarné par l'État central, l'État a souvent fait suivre la décentralisation d'une compétence par son évaluation. On retrouve là, une logique inverse par rapport aux États-Unis, où l'histoire de la construction politique situe l'intérêt général au niveau local des États fédérés, ce qui explique le développement de l'évaluation au niveau fédéral par nécessité de légitimation de son intervention⁴⁷⁹. L'évaluation a été aussi impulsée par une dynamique européenne dans le cadre de politiques contractuelles liant les Conseils Régionaux et l'Union Européenne. En effet, par le biais des procédés de cofinancement, la Commission européenne, par la réforme des fonds structurels (FSE et le Feder) de 1988, a organisé principalement des relations avec le local autour de politiques contractuelles voire expérimentales⁴⁸⁰.

Ensuite, elle a peu à peu été pratiquée à l'initiative des collectivités locales et porte sur des politiques relevant de leurs propres compétences. Comme le remarque P. Duran et E. Monnier, la décentralisation et l'évaluation s'inscrivent dans des logiques communes, d'une part de technicité, c'est-à-dire qu'elles cherchent à améliorer le rendement du système de décision publique, et d'autre part d'une dimension démocratique dans la volonté qu'elles manifestent de se rapprocher des citoyens⁴⁸¹. L'évaluation a permis de donner à voir ces collectivités en action et d'accéder à une logique d'autolégitimation de leur intervention et de leurs compétences⁴⁸². Selon V. Chanut, même en Région, cette pratique reste malgré tout vulnérable, en témoigne le positionnement incertain des cellules d'évaluation et le fort turn-over des agents s'en occupant⁴⁸³.

⁴⁷⁸ Eric MONNIER et Fabrice HENARD, « L'Évaluation des politiques publiques dans les collectivités territoriales : exercices imposés et initiatives propres », *Pouvoirs locaux* (novembre 2000).

⁴⁷⁹ Jean-Pierre NIOCHE, *op. cit.* p. 44.

⁴⁸⁰ Alain FAURE et Andrew SMITH, « L'évaluation objet de savoir et enjeu de pouvoir, le cas des politiques publiques locales », *Sciences de la Société* 1994.32.101-111, p.105.

⁴⁸¹ Patrice DURAN et Eric MONNIER, *op. cit.* p. 237.

⁴⁸² Clément LACOUETTE-FOUGERE et Pierre LASCOURMES, *op. cit.* p. 76.

⁴⁸³ Véronique CHANUT, « Contes et mécomptes d'une connaissance ordinaire », (2003) 11-57 *Pouvoirs locaux* 99-105, p.99.

Le champ de l'offre d'évaluation en France se structure principalement autour de moyens ou petits cabinets⁴⁸⁴. Selon l'étude de N. Matyjasik en 2006, le marché de l'évaluation est constitué à plus de 90% par des cabinets de conseils et des consultants indépendants dont la majorité (61% soit 87 sur 138) ont un effectif compris entre 1 et 10 salariés, et dont 62% ont vu le jour à partir des années 1990⁴⁸⁵. Néanmoins, les frontières ont été poreuses entre quelques milieux académiques de politologues et de sociologues et les cabinets des années quatre-vingt-dix. En témoigne le cas du laboratoire CEOPS (Conception d'Évaluation pour les Organisations et les Politiques publiques) créé à Lyon en 1988 par le ministère de l'équipement au sein de l'École Nationale des Travaux Publics de l'État (ENTPE) qui réunit une équipe de chercheurs, dont V. Spenlehauer et P. Warin, sous la direction d'E. Monnier. Par leur participation à la rédaction des cahiers MEANS en 1999 établissant des guides de pratiques d'évaluation pour la Commission européenne, une partie des membres de ce laboratoire fonde le Centre Européen d'Expertise en Evaluation, un des premiers cabinets d'évaluation en France⁴⁸⁶. Ce choix, comme l'explique un des membres fondateurs dans un entretien avec N. Matyjasik, est issu en partie d'une incompatibilité des activités de recherche et d'évaluation. « *Ce n'était pas satisfaisant, car l'évaluation ne donne pas lieu à publication... Pour un chercheur, c'est très frustrant, cela n'a aucun intérêt académique. Pour un ministère, c'est frustrant (un travail académique) quand on attend des réponses concrètes à des questions* »⁴⁸⁷.

La préférence pour le modèle de l'évaluation pluraliste⁴⁸⁸ en France, à partir des années 1990, a ouvert la voie à l'intervention d'acteurs « non scientifiques », c'est-à-dire dont l'activité principale n'est pas la recherche, et qui ne fondent pas leur légitimité sur la maîtrise d'une méthode scientifique de démonstration en particulier⁴⁸⁹. En effet, comme l'ont souligné E. Guba et Y. Lincoln à propos de la 4^{ème} génération d'évaluation correspondant au modèle pluraliste français, l'évaluateur devient un « médiateur », facilitant l'arbitrage entre les intérêts des parties prenantes, il soutient ainsi sa légitimité non plus sur l'application d'une méthode scientifique mais sur la compréhension des différents enjeux évaluatifs et sa capacité à faire émerger un consensus⁴⁹⁰. Cette démarche s'intéressant particulièrement au processus de formulation des politiques publiques et de leur mise en œuvre, ouvre la pratique de l'évaluation à des personnes formées aux sciences politiques et aux techniques d'apprentissage organisationnel⁴⁹¹.

Les scientifiques sont restés à l'écart de ce développement. Alors que l'orientation empiriste et finalisée de la recherche en sciences sociales aux États-Unis a été un moteur pour le développement de l'évaluation dans ce pays, l'orientation française à l'écart de l'action publique a freiné le développement. Pour reprendre la distinction d'H. D. Lasswell, l'analyse américaine des politiques publiques s'est davantage inscrite

⁴⁸⁴ Jean-Claude BARBIER, *op. cit.* p.39. Ludovic MEASSON et Nicolas MATYJASIK, « L'évaluation entre opportunité et effet », dans Alain FAURE et ASSOCIATION FRANÇAISE DE SCIENCE POLITIQUE (dir.), *Les politiques publiques à l'épreuve de l'action locale: critiques de la territorialisation*, coll. Questions contemporaines, L'Harmattan, 2007 p. 170- 176.

⁴⁸⁵ Nicolas MATYJASIK, *L'évaluation des politiques publiques dans une France décentralisée, institutions, marché et professionnels*, Université Montesquieu-Bordeaux IV/Sciences-po Bordeaux, 2010, p. 188-196.

⁴⁸⁶ Aussi appelé Euréval et au sein duquel nous avons conduit notre contrat Cifre Nicolas MATYJASIK, *op. cit.* p. 180.

⁴⁸⁷ Propos reportés d'un entretien mené dans le cadre du travail de thèse de Nicolas MATYJASIK, *Ibid.*

⁴⁸⁸ L'évaluation pluraliste est un processus d'évaluation centré sur l'association des principales parties prenantes de la politique évaluée: décideurs, opérateurs et représentants de la société civile. Au-delà d'estimer l'effet de l'intervention, son ambition est aussi de développer une communication entre ces parties et leur accord sur des constats communs Gaëlle BARON et Eric MONNIER, « Une approche pluraliste et participative: Coproduire l'évaluation avec la société civile », *Informations sociales* 2003.p. 110.

⁴⁸⁹ Paul SABATIER, « The political context of evaluation research, an advocacy coalition perspective », dans Marie-Christine KESSLER, Pierre LASCOUMES et Jean-Claude THOENIG, *Evaluation des politiques publiques*, coll. Collection Logiques politiques, Paris, Harmattan, 1998, p. 129- 146 à la page 26.

⁴⁹⁰ E GUBA et Y LINCOLN, *op. cit.*.

⁴⁹¹ Randall B. RIPLEY, *op. cit.* p. 179.

historiquement dans le « *knowledge in the policy process* » alors que la tradition française était une inscription dans le « *knowledge of the policy process* »⁴⁹². En effet, le monopole de l'expertise d'État par les administrateurs de l'INSEE et les grands corps décrit ci-dessus a contribué à maintenir l'Université à l'extérieur des réalités politiques et à laisser les chercheurs français avec des opportunités limitées d'accès au « terrain » de l'action publique, et donc sans expérience dans le domaine⁴⁹³.

En somme, il semble que la configuration historique de l'interaction des sciences sociales et de l'action publique n'ait pas été favorable au développement de l'évaluation au sein de l'administration centrale française. Ce bref retour historique témoigne d'un développement cloisonné du recours à l'expertise, constitué d'une part, autour d'une relation privilégiée avec les Grands Corps et l'INSEE, et d'autre part de l'évaluation des politiques publiques, pratiquée principalement par des acteurs n'étant pas des chercheurs, au sein des collectivités territoriales.

L'expérimentation aléatoire, de prime abord, ne semble pas apparaître dans le paysage français de l'expertise et de l'évaluation des politiques publiques avant les années 2000. L'expérimentation normative apparaît de façon marginale, mais ses modalités d'évaluation ne sont pas précisées. Néanmoins, en étudiant de plus près les débats ayant jalonné et structuré les milieux de l'évaluation en France, la méthode expérimentale par assignation aléatoire se retrouve à plusieurs reprises discutée. Au cours des années 2000, elle apparaît peu à peu comme l'instrument suscitant l'intérêt des statisticiens pour l'évaluation des politiques publiques, et ouvre alors de nouvelles possibilités pour le développement de cette démarche. L'adoption de cette méthode, ne tient donc pas uniquement à la diffusion internationale d'un instrument d'action publique et celle-ci émerge dans les débats français uniquement au cours de la dernière décennie.

1.1.2. L'expérimentation en France avant la « méthode expérimentale »

La méthode expérimentale regroupant intervention politique innovante et méthode d'évaluation expérimentale par assignation aléatoire a connu un essor sans précédent en France à la fin des années 2000. Présentée comme une innovation à cette époque, son pendant politique est néanmoins présent dans des démarches d'innovation politique dès les années 1970, quant à son pendant scientifique il été discuté dès les débuts de l'évaluation en France, dans les années 1980. L'intérêt pour cette méthode étant resté dormant pendant une vingtaine d'années, il a été ravivé par l'approfondissement du mouvement décentralisateur, et le renouveau d'un intérêt académique pour cette méthode d'évaluation scientifique.

1.1.2.1. L'expérimentation politique, emploi parcimonieux et dynamique de décentralisation

La démarche expérimentale en action publique en France n'est pas récente. Son développement est néanmoins resté assez restreint et avec la particularité de comporter une exigence d'évaluation sans pour autant en préciser la méthode. L'expérimentation juridique a pris deux formes en France.

⁴⁹² H.D. Lasswell cité par Patrice DURAN, « Genèse de l'analyse des politiques publiques », dans Pauline RAVINET et Pierre MULLER, *dictionnaire des politiques publiques*, Presses de Science Po, Paris, 2006.

⁴⁹³ Alain FAURE et Andrew SMITH, *op. cit.* p. 103. D'autres soulignent aussi la faible sollicitation des chercheurs en science sociale dans le processus d'institutionnalisation de l'évaluation Spenlehauer, « Intelligence gouvernementale et sciences sociales », p.127.

Tout d'abord, l'expérimentation dite normative qui « *consiste à tester, pendant une période donnée, une nouvelle règle de droit: le législateur ou le pouvoir réglementaire statuent à titre provisoire, sous bénéfice d'inventaire, au vu des résultats produits, la règle sera pérennisée dans les mêmes termes, ou modifiée, ou abandonnée.* »⁴⁹⁴. Dans ce cadre, le premier exemple d'expérimentation à l'échelle nationale, est la loi de 1975 sur l'interruption volontaire de grossesse, qui avait été promulguée « *à l'essai* » pour une durée de cinq ans, avec une « *clause de rendez-vous* ». Dans cette même logique, on retrouve la loi du 1^{er} décembre 1988, instaurant le RMI jusqu'au 30 juin 1992, soumis à une évaluation, date à laquelle le gouvernement devait déposer un projet de loi⁴⁹⁵. La démarche visait plus à faciliter l'adoption de cette loi controversée plutôt que de juger de son efficacité⁴⁹⁶. J.-P. Nioche, dans un article de 1982, relève la démarche expérimentale des politiques de sécurité routière, ainsi que l'initiative en 1976 du Ministère du Travail, de mener des expériences pour adapter le projet de loi sur le bilan social dans les entreprises ; initiatives qui n'ont pas survécu au changement de ministre.

Une autre logique expérimentale est historiquement inscrite dans la logique de décentralisation. Elle désigne des actions ponctuelles, autorisées par l'État dans le cadre de dérogation à des lois générales. On retrouve cette logique dès 1962, pour apprécier la pertinence d'une mesure de déconcentration administrative qu'on applique à cinq départements et deux régions. La conduite de ce type d'innovation a ensuite été encouragée par les fonds européens, comme le FSE (Fonds Social Européen) qui, en 1994, réservait 5% de son budget à des programmes innovants, ainsi que le programme Equal de 2002 à 2008, qui avait vocation à financer les innovations sociales relatives au marché de l'emploi et à l'inclusion sociale⁴⁹⁷. Ces politiques contractuelles sont considérées comme des expérimentations. Cependant, pour reprendre les termes d'A. Faure et A. Smith, « *le passage des expérimentations à l'analyse des résultats comporte bien des obstacles* », au sens où ces études portent principalement sur les réalisations du projet et non sur les résultats, et qu'elles sont difficiles à analyser et rarement prises en compte par les commanditaires⁴⁹⁸. Ce développement est resté très anecdotique. La démarche expérimentale est néanmoins mentionnée dans le rapport Deleau sur l'évaluation de 1986, elle est envisagée comme « *une démarche pragmatique (qui) permettrait d'éviter des déboires coûteux* »⁴⁹⁹.

⁴⁹⁴ Gwénaële CALVES, *Histoire et perspectives du cadre juridique des expérimentations législatives et réglementaires en France*, Paris, Dares, Colloque sur les expérimentations pour les politiques publiques de l'emploi et de la formation, 22 mai 2008, p. 95-106, p. 95.

⁴⁹⁵ Bernard GOMEL et Evelyne SERVERIN, « Expérimenter pour décider? Le RSA en débat », *Document de travail Centre d'étude de l'Emploi* 2009.119, p.9.

⁴⁹⁶ Jean-Pierre NIOCHE, *op. cit.* p.38.

⁴⁹⁷ Olivier NOBLECOURT et Frédéric VALLETOUX, *L'expérimentation sociale locale en France: bilan, enjeux, perspectives*, Rapport de mission 2009-2010, rapport non publié, p. 18; COMITE ECONOMIQUE ET SOCIAL EUROPEEN, *Avis du Comité économique et social européen sur «Comment l'expérimentation sociale peut elle servir en Europe à l'élaboration des politiques publiques d'inclusion active»*, (2009) (2009/C 100/13).

⁴⁹⁸ Alain FAURE et Andrew SMITH, *op. cit.* p. 106.

⁴⁹⁹ Ce rapport identifie les lacunes de la France en évaluation la première lacune est « *D'abord celui du recours à l'expérimentation, très peu développé en France. L'expérimentation, c'est-à-dire l'application d'une politique à un groupe, de taille restreinte, mais représentatif, avec mesure des résultats avant généralisation éventuelle, ne saurait constituer une panacée. Mais, sans nul doute, un usage raisonné de cette démarche pragmatique permettrait d'éviter des déboires coûteux* » Michel DELEAU, *Evaluer les politiques publiques: méthodes, déontologie, organisation: rapport du Groupe de travail « Méthodes d'évaluation des politiques publiques »*, coll. Plan / Commissariat général du plan, n°13, Paris, Documentation française, 1986, p. 8.

Ces exceptions à la constitution sont entérinées par la modification constitutionnelle de 2003. Une première jurisprudence de 1993 avait amorcé ce travail. Une décision du Conseil Constitutionnel du 28 juillet 1993, autorise la pratique expérimentale en ayant pour critère la durée, qui doit être définie, la conception autour d'une généralisation éventuelle, et la définition des procédures d'évaluation⁵⁰⁰. Le 16 janvier 2001, le Député Pierre Méhaignerie, président de la commission des finances, fait voter, à une très grande majorité, une proposition de loi demandant l'introduction dans la constitution du droit à l'expérimentation⁵⁰¹. Le nouvel article 37-1 confirme la possibilité d'introduire dans la loi des dispositions à caractère expérimental dérogeant au principe d'égalité, par exemple pour réaliser un nouveau transfert de compétences.

L'acte 2 de la Décentralisation ouvre, par le quatrième aliéna de l'article 72 de la Constitution, dans sa rédaction issue de la loi constitutionnelle relative à l'organisation décentralisée de la République, « la possibilité pour les collectivités territoriales et leurs groupements, lorsque selon le cas la loi ou le règlement l'a prévu, de déroger, à titre expérimental et pour un objet et une durée limités, aux dispositions législatives ou réglementaires qui régissent l'exercice de leurs compétences »⁵⁰². En terme d'évaluation, la loi prévoit qu' : « Avant l'expiration de la durée fixée pour l'expérimentation, le gouvernement transmet au Parlement, aux fins d'évaluation, un rapport assorti des observations des collectivités territoriales qui ont participé à l'expérimentation. Ce rapport expose les effets des mesures prises par ces collectivités en ce qui concerne notamment le coût et la qualité des services rendus aux usagers, l'organisation des collectivités territoriales et des services de l'État ainsi que leurs incidences financières et fiscales »⁵⁰³. »

L'expérimentation est perçue par le député ayant proposé cette loi comme « une nouvelle forme de gouvernance » permettant de faire mieux et moins de lois, de faciliter l'adoption de réformes difficiles en incarnant les aspects positifs de celles-ci, de laisser une marge d'initiative aux collectivités, et enfin de porter la question de l'évaluation, sans plus de précisions méthodologiques, au-devant de la scène⁵⁰⁴.

Selon nos interlocuteurs, la démarche expérimentale qui consiste à mettre en place de manière temporaire un dispositif afin d'en apprécier les effets est assez populaire, d'autant plus que sur des sujets polémiques elle permet de laisser le temps de calmer le débat. Cependant, malgré l'ouverture du droit à l'expérimentation locale, cette forme d'intervention publique n'est pas mobilisée par les collectivités⁵⁰⁵. Le renouveau de l'intérêt des chercheurs pour la méthode expérimentale par assignation aléatoire dans les années 2000 est un élément déterminant dans la promotion de l'expérimentation en action publique.

1.1.2.2. L'expérimentation aléatoire, un idéal statistique « inapplicable » en France

Si l'utilisation de la méthode expérimentale par assignation aléatoire en France est un phénomène récent, son évocation dans les débats scientifiques autour de l'évaluation des politiques publiques remonte au début des débats français en la matière. Après avoir été

⁵⁰⁰ AGENCE NOUVELLE DES SOLIDARITES ACTIVES, « Revue succincte d'activité », 2006.2.

⁵⁰¹ Pierre MEHAIGNERIE, « Décentralisation, expérimentation et évaluation », (2003) 11-57 *Pouvoirs locaux* 113-116.

⁵⁰² LOI organique n° 2003-704 du 1er août 2003 relative à l'expérimentation par les collectivités territoriales (1), (2003).

⁵⁰³ *Ibid.*

⁵⁰⁴ Pierre MEHAIGNERIE, *op. cit.*

⁵⁰⁵ COMMISSION FAMILLES, VULNERABILITE, PAUVRETE, *Au possible, nous sommes tenus. La nouvelle équation sociale, 15 résolutions pour combattre la pauvreté des enfants*, 2005.

peu à peu délaissée, elle connaît un regain d'intérêt dans les années 2000, suite notamment à son utilisation croissante dans les travaux internationaux en économie. Cette partie ne prétend pas retracer de manière exhaustive toutes les racines de l'expérimentation aléatoire en France. Elle s'appuie sur les entretiens menés auprès des praticiens de cette méthode, ou d'observateurs de longue date du champ de l'évaluation en France, à partir d'un double questionnement sur ce qui les avait amenés à s'intéresser à cette méthode, et sur l'historique de son développement. Ce paragraphe est donc teinté de la subjectivité des acteurs rencontrés dans notre étude, ainsi que de celle des documents collectés faisant référence à cette méthode

Cette méthode d'évaluation est présente à l'origine des débats français en évaluation des politiques publiques. Elle est déjà mentionnée par J.-P. Nioche dans un des premiers textes sur l'évaluation en France en 1982. Ce professeur à HEC (École des Hautes Études Commerciales) inscrit l'expérimentation dans une démarche de recherche évaluative de mise à l'épreuve empirique d'une hypothèse théorique, en se référant aux travaux de D. Campbell⁵⁰⁶. Il identifie « l'expérimentation », ici entendue comme méthode scientifique, comme un des dispositifs de recherche - avec les études économétriques et les études de cas - pouvant remplir cette tâche. Il souligne néanmoins que « *la procédure expérimentale stricte, qui prévoit la répartition au hasard des individus entre un groupe expérimental, soumis à la politique que l'on veut évaluer, et un groupe de contrôle, qui n'y est pas soumis, soulève de graves problèmes éthiques et politiques si on veut l'utiliser dans un certain nombre de domaines « sensibles »* », amenant de fait au développement de méthodes quasi expérimentales⁵⁰⁷.

Lors du premier grand colloque français sur l'évaluation en 1983, cette méthode et son application américaine ont été aussi débattues. F. Levy, de l'*Urban Institute*, la présente comme une méthode pouvant produire les « *meilleurs résultats d'évaluation* » lorsqu'elle est bien structurée. Il explique son lien avec les normes du champ scientifique en ces termes : « *Au plan méthodologique, ce qui a trait à l'expérimentation, c'est-à-dire les techniques d'expériences aléatoires, constitue au fond l'élément le plus original de l'évaluation. La méthode de l'évaluation est celle que l'on emploie dans les travaux de recherche pour valider une théorie scientifique. Elle doit avoir un caractère systématique et rigoureux* »⁵⁰⁸. Il revient alors sur les risques qu'elle engendre, notamment liés à l'interprétation des échantillons tronqués, il acte le manque d'enthousiasme des politiciens pour cette méthode à cause de son coût, la restriction des droits des citoyens qu'elle engendre, et le fait qu'elle retarde l'action politique dans l'attente des résultats⁵⁰⁹. L'expérimentation aléatoire est très discutée lors de ce colloque du début des années 1980. Elle est présentée par R. Anglemar comme étant « *le plan de recherche le plus élégant* » bien qu'il circoncrive précisément son emploi, et D. Schwartz présente son utilisation dans le champ de l'épidémiologie⁵¹⁰. Lors de ce colloque, la discussion conclut à la reconnaissance de l'intérêt pour l'expérimentation, mais souligne la difficulté de son acceptation par les instances politiques et notamment le Conseil d'État⁵¹¹.

⁵⁰⁶ Jean-Pierre NIOCHE, *op. cit.* p. 36.

⁵⁰⁷ *Ibid.* p. 46.

⁵⁰⁸ Jean-Pierre NIOCHE et Robert POINSARD, *L'Évaluation des politiques publiques*, Paris, Economica, 1984, p. 295.

⁵⁰⁹ Frank LEVY, « L'évaluation des politiques publiques aux États-Unis », dans Jean-Pierre NIOCHE et Robert POINSARD, *L'Évaluation des politiques publiques*, Paris, Economica, 1984, p. 32-33.

⁵¹⁰ Voir p 83-84 pour la présentation de la méthode par R. Anglemar et p 93-105 pour la présentation des essais cliniques par D. Schwartz dans Jean-Pierre NIOCHE et Robert POINSARD, *op. cit.*

⁵¹¹ Jean-Pierre NIOCHE et Robert POINSARD, *op. cit.* p. 299.

Cette méthode est peu à peu mise de côté par les orientations démocratiques et pluralistes de l'évaluation française. Elle est d'ailleurs mobilisée en creux pour légitimer ces formes d'évaluations plus participatives. Dans un ouvrage de 1992, E. Monnier leur consacre un chapitre et présente des cas d'utilisation. Il évoque la déception des décideurs et des universitaires vis-à-vis de l'utilisation de cette méthode aux États-Unis et revient en particulier sur l'emblématique expérimentation d'impôt négatif du New Jersey (*New Jersey Negative Income Tax*), l'une des toutes premières expérimentations par assignation aléatoire de grande ampleur aux États-Unis effectuée en 1968. Il relève plusieurs difficultés de validité interne, à savoir le biais des résultats selon le taux d'abandon, de 22% pour cette expérimentation⁵¹². Il présente aussi, les difficultés de validité externe, comme le biais d'auto sélection⁵¹³, l'effet Hawthorne, ou les spécificités liées au contexte local de l'expérimentation. Il note, au regard des expériences américaines, l'incompatibilité de cette méthode avec la temporalité de la décision politique ainsi que les modalités de la prise de décision correspondant davantage à une situation de compromis que d'optimisation rationnelle. Il rend compte des débats de la communauté scientifique américaine. Les exigences de validité interne entrant en contradiction avec celles de la validité externe; en d'autres termes, plus la situation expérimentée est artificielle, plus elle s'éloignera d'un contexte d'application banale non expérimentale⁵¹⁴. Il souligne que la plupart des évaluateurs américains affectent l'utilisation de la méthode aléatoire à un « *nombre extrêmement restreint de cas* »⁵¹⁵. Son bilan a pour but explicite d'éviter la transposition en France de ces expérimentations « *à l'américaine* » et de promouvoir à la place une évaluation pluraliste⁵¹⁶.

Jusque dans les années 2000, cette méthode d'évaluation continue à être présentée comme un des outils de l'évaluation des politiques publiques, mais reste non pratiquée. Elle est conduite dans d'autres milieux comme celui de la sécurité alimentaire ou dans le cadre des essais cliniques, mais ces mondes communiquent très peu avec le monde de l'évaluation des politiques publiques⁵¹⁷. Elle reste néanmoins une méthodologie dormante en France puisqu'elle est perçue comme « *le Graal* », pour reprendre les termes de nos interlocuteurs, résolvant les problèmes de sélection de groupes de comparaisons pour les statisticiens s'intéressant à la causalité, mais elle se heurte à l'interdiction imposée par la norme juridique. Un statisticien du CREST se souvient de l'opposition de la culture française à cette méthode : « *l'inégalité d'accès à des mesures, à des dispositifs était considérée comme le mal absolu* ».

L'intérêt pour cette méthode s'inscrit dans un mouvement de mathématisation de l'économie, tendance particulièrement présente chez les économistes français, notamment à travers la statistique. Cette dernière est le médium qui permet à l'économie de devenir un savoir appliqué⁵¹⁸. Tous les chercheurs rencontrés mobilisant la méthode expérimentale se définissent comme économistes appliqués⁵¹⁹. Cette méthode s'inscrit

⁵¹² Eric MONNIER, *Evaluations de l'action des pouvoirs publics*, op. cit. p. 33.

⁵¹³ les gens acceptant d'entrer dans l'expérimentation peuvent avoir des caractéristiques différentes que le reste de la population

⁵¹⁴ Eric MONNIER, *Evaluations de l'action des pouvoirs publics*, op. cit. p. 46.

⁵¹⁵ Ibid. p. 47.

⁵¹⁶ Ibid. p. 50.

⁵¹⁷ Pour avoir un éclairage du développement des essais cliniques en France voir l'article Nicolas DODIER et Janine BARBOT, « Le temps des tensions épistémiques: Le développement des essais thérapeutiques dans le cadre du sida », (2000) 41-1 *Revue française de sociologie* 79-118.

⁵¹⁸ Delphine DULONG, « Quand l'économie devient politique. La conversion de la compétence économique en compétence politique sous la Ve République », op. cit. p. 115.

⁵¹⁹ Le discours d'un de nos interlocuteurs statisticien est un exemple représentatif de cette philosophie « *Moi j'ai décidé de faire de l'économie justement pour la capacité que peu avoir la science à comprendre et éclairer le fonctionnement de notre vie de tous les jours le comportement des individus* ».

dans la recherche des économistes du *ceteris paribus* (« toute chose égale par ailleurs »), c'est-à-dire de l'établissement d'une causalité par l'estimation de l'impact de la manipulation d'une variable en comparant deux groupes exactement identiques, seulement différenciés par la présence ou non de cette variable. Mais cette reproduction d'un groupe à l'identique étant impossible, les chercheurs doivent mobiliser des outils pour reconstruire ce qu'on appelle le scénario contrefactuel. Un statisticien du centre d'étude pour l'emploi (CEE) se remémore : « *construire des groupes témoins a posteriori au CEE, on avait essayé de faire ça, en comparant des sites, en essayant de les rendre toutes choses égales par ailleurs sur les aides aux chômeurs, c'était très compliqué, je me souviens, on s'arrachait les cheveux, donc on faisait de l'économétrie à tout va pour essayer...mais il y avait toujours ces variables cachées* ». Les scientifiques étudient donc la possibilité de conduire cette méthode dans le cadre juridique français, comme en témoigne une étude de faisabilité d'une expérimentation contrôlée pour la délégation interministérielle à l'insertion des jeunes, parue en 1998⁵²⁰.

Ces méthodes restent, malgré leur faible utilisation, présentes dans la culture des statisticiens français et des économistes appliqués. Elles sont enseignées dans les écoles de statistiques, via la présentation de grandes expérimentations américaines. Selon un interlocuteur, elles ont « *marqué les esprits scientifiquement (...) ce sont des références absolues* », notamment par leur démonstration de la causalité. Certains de ces chercheurs entrent aussi à son contact par des séjours universitaires à l'étranger. Cette immersion renforce leur motivation. Par exemple, un économiste français qui a suivi aux États-Unis les cours de D. Rubin au milieu des années quatre-vingt-dix explique « *je suis rentré en France après et vraiment avec une seule envie, c'était de faire ça et donc c'est ce que j'ai essayé de faire de multiples façons* ». Un autre statisticien après avoir suivi les cours de J. Angrist à Harvard explique que son discours sur la causalité en statistique se positionne en porte-à-faux avec l'enseignement français. Alors que ce dernier présentait la statistique uniquement comme un outil de démonstration de la corrélation, ces économistes américains disaient « *on peut avoir une inférence causale crédible, mais pour ça il faut soit construire son expérience soit bénéficier de quelque chose qui y ressemble* »⁵²¹.

Les économistes français souhaitant adopter une approche expérimentale se replient sur des quasi-expériences qui s'appuient sur des données « naturelles » créées par des configurations politiques particulières. On note à ce sujet les travaux de T. Piketty sur l'impact des incitations à l'emploi par la baisse des cotisations sociales patronales en 1993⁵²². Aussi, D. Fougère étudie les effets d'une réforme de prestations sociales au Québec en 1989⁵²³. Ils pratiquent des régressions par discontinuité pour simuler artificiellement des situations expérimentales⁵²⁴. Celles-ci sont envisagées comme des « *substituts avantageux* » puisqu'elles n'ont pas le coût politique de l'expérimentation contrôlée, néanmoins, leur biais statistique quant à la constitution du groupe de contrôle

⁵²⁰ Fougère, D., S. Lau, T. Magnac et M. Visser (1998) : « Etude de faisabilité d'une expérimentation contrôlée pour l'évaluation d'une mesure publique d'aide à l'emploi », Rapport final pour la Convention 97-000 20 passée avec la Délégation Interministérielle à l'Insertion des Jeunes, CREST-INSEE, Paris.

⁵²¹ Source : entretien auprès d'un statisticien

⁵²² T. Piketty, 1998 « l'impact des incitations financières au travail sur les comportements individuels : une estimation pour le cas français », *Economie et Prévision*, n° 132-133.

⁵²³ Fortin, B., D. Fougère et G. Lacroix (1999) : « Hausses des barèmes et sorties de l'aide sociale. Les résultats d'une expérience naturelle au Canada », *Revue Economique*, vol. 50 (3), p. 451-464.

⁵²⁴ Cette méthode consiste à utiliser un seuil d'accès à la prestation créant une discontinuité de traitement parmi une population aux caractéristiques très proches, lorsque l'on prend les personnes les plus proches de ce seuil. Par exemple l'accès à une aide selon un seuil de revenu. Cela permet de comparer une population avec et sans traitement aux caractéristiques quasiment similaires. Une analyse de régression est appliquée pour comparer l'évolution des individus au dessus et en dessous du seuil. SOCIÉTÉ FRANÇAISE D'ÉVALUATION, *L'évaluation des impacts des programmes et services publics*, Cahier de la SFE n°6, septembre 2011, p. 28.

sont de plus en plus mis à jour⁵²⁵. Ces statisticiens sont partagés entre ceux qui souhaitent conduire des expérimentations aléatoires, mais qui restent isolés, et les autres membres du laboratoire qui s'inscrivent davantage dans la tradition française de dextérité économétrique. La méthode expérimentale aléatoire est vue comme une méthode simpliste ne mettant pas en valeur les compétences économétriques pointues⁵²⁶. Cependant, comme nous l'avons évoqué au premier chapitre, au courant des années 2000, la méthode expérimentale devient peu à peu, dans le *champ d'action stratégique* international de la science économique appliquée, une ressource dont la maîtrise et la conduite sont un atout notoire pour sécuriser une place de *titulaire*.

On retrouve des traces du renouveau de l'intérêt scientifique pour la méthode expérimentale par assignation aléatoire dès les années 2000, par le biais notamment de mobilisation de données ou de travaux étrangers. Plusieurs chercheurs rencontrés ont évoqué l'importance dans ce développement de l'article de D. Fougère (laboratoire Crest-Insee) de 2000, intitulé « *Expérimenter pour évaluer les politiques d'aide à l'emploi: les exemples anglo-saxons et Nord européens* ». Cet article présente les fondements philosophiques et scientifiques de « *l'évaluation par expérimentation contrôlée* », et ses capacités à corriger les biais statistiques induits par les méthodes quasi expérimentales, comme particulièrement l'éventuelle endogénéité de la variable de contrôle et le biais de sélection⁵²⁷. Il revient sur les expériences américaines dans le domaine des politiques d'aides ou d'incitation au retour à l'emploi (*Job training Partnership Act* par exemple) en étudiant les arguments de réticence vis-à-vis de cette méthode, et les blocages institutionnels qu'elle génère auprès des opérateurs⁵²⁸. Il finit enfin, sur les expériences européennes (*Restart* en Grande-Bretagne en 1989) et conclut que les expérimentations sur les programmes d'intensification des aides à la recherche d'emploi sont les plus aptes à produire des résultats robustes au vu de la plus grande aptitude à les « *contrôler* ».

E. Maurin, du laboratoire Crest-Insee, dans un article de 2002, mobilisait déjà des résultats de l'expérimentation aléatoire américaine « *Link up* » pour discuter de l'efficacité des dispositifs de lutte contre les trappes à chômage⁵²⁹. Il met en lumière dans cet article les résultats de « *l'expérience contrôlée* », très éloignés de la simulation économétrique fondée selon lui sur de fausses hypothèses⁵³⁰. L'avantage de cette méthode est : « *qu'il n'est nul besoin de faire l'hypothèse (...) pour évaluer les effets d'une réforme* », mais il estime cependant qu'elle ne peut éclairer les comportements que pour une zone précise du territoire⁵³¹.

La modification constitutionnelle de 2004 lève alors les barrières normatives, principaux obstacles à la conduite de cette méthode, elle ouvre de nouveaux horizons aux chercheurs. Un ancien directeur de la DARES se rappelle : « *Alors là, les statisticiens qui rongeaient leur frein comme moi depuis des années, ou depuis moins longtemps pour*

⁵²⁵ Denis FOUGERE, « Expérimenter pour évaluer les politiques d'aide à l'emploi: les exemples anglo-saxons et nord-européens. », *Revue Française des Affaires Sociales* 2000, 10.

⁵²⁶ « les autres membres du Crest ne sont pas finalement si convaincus que ça de l'expérimentation, ils le deviennent parce qu'ils voient qu'il y a beaucoup de papiers qui sont publiés. C'est pas le truc français ça, parce que ça repose pas assez, à mon avis, sur la dextérité mathématique, la dextérité économétrique ou l'astuce de la quasi expérimentation » Ce que souligne ici ce statisticien est la tradition particulièrement française de mathématisation de l'économie par le développement de méthodes quasi-expérimentales ou de modélisations. Ses représentants n'ont donc pas trouvé un intérêt à la méthode expérimentale qui ne reposait pas sur des calculs élaborés.

⁵²⁷ Denis FOUGÈRE, *op. cit.* p. 11.

⁵²⁸ Il revient notamment sur le coût et le temps consacré par le MDRC (cabinet privé américain en charge de l'évaluation) en éléments de communication ou d'information autour du protocole Denis FOUGERE, *op. cit.* p. 3.

⁵²⁹ Alisa LEWIN et Eric MAURIN, « Quels effets attendre de réformes permettant de cumuler emploi et aide sociale? Une évaluation sur données expérimentales américaines », (2002) 53-6 *Revue économique* 1357-1368.

⁵³⁰ *Ibid.* p.1366.

⁵³¹ *Ibid.*

les plus jeunes, ils s'y sont engouffrés ». Mais, malgré l'ouverture normative, les chercheurs souhaitant la pratiquer réalisent que son emploi nécessite au préalable un changement de culture administrative et politique pour accepter le tirage au sort expérimental. Une poignée de statisticiens du Crest amorcent un *travail institutionnel* de diffusion cognitive des apports de cette méthode. Par exemple, en avril 2004, lors des ateliers de M2S3 (Modélisation et Méthodes Statistiques en Science Sociales) du Lasmass⁵³² sur « *les méthodes d'évaluation statistiques des politiques publiques* », M. Gurgand du laboratoire Delta et Crest-Insee présentait les travaux d'E. Duflo sur la construction d'écoles en Indonésie, ainsi que l'expérimentation aléatoire canadienne du *Self sufficiency project*⁵³³.

En parallèle, dans les milieux de l'économie du développement en France, la méthode expérimentale, et particulièrement les travaux d'E. Duflo, commencent à être discutés. En novembre 2004, à l'occasion de la deuxième conférence AFD/EUDN (Agence Française de Développement/European Development Research Network) sur le thème « *Aide au développement : pourquoi et comment-quelles stratégies pour quelle efficacité?* » Esther Duflo du Massachusetts Institute of Technology et Poverty action lab, présente une communication⁵³⁴. Ancienne élève de l'Ecole normale supérieure, et détentrice d'un DEA en économie du centre Delta, elle garde des liens étroits avec les économistes parisiens de ce laboratoire, notamment avec Thomas Piketty et François Bourguignon. Elle participe avec ce dernier au même groupe de réflexion « *the evaluation gap working group* » convoqué par le *Center for global Development* la même année, déjà évoqué dans notre premier chapitre⁵³⁵. Les travaux de ce groupe sont orientés autour de la recherche des moyens de promotion de la conduite d'évaluation d'impact pour les programmes de développement, dont la méthode expérimentale par assignation aléatoire. François Bourguignon, alors économiste en chef et vice-président sénior pour l'économie du développement à la Banque mondiale, se présente comme un promoteur très actif de ces méthodes auprès de cette instance à cette période⁵³⁶.

Cette intervention, « *Évaluer l'impact des programmes d'aide au développement : le rôle des évaluations par assignation aléatoire* », a donné lieu à une publication en 2005 dans la Revue d'économie du développement. Elle prend la forme d'un plaidoyer pour étendre l'utilisation de la méthode expérimentale par assignation aléatoire. Cet argumentaire sera ensuite repris par les promoteurs de cette méthode en France. E. Duflo cadre le débat autour de l'enjeu central, selon elle, des méthodologies d'évaluation d'impact, qui est de constituer un groupe de comparaison crédible, un contrefactuel - groupe n'ayant pas bénéficié de l'intervention publique - mais identique en tous points au groupe en ayant bénéficié⁵³⁷. Elle présente dans un premier temps l'apport scientifique de cette méthode. Après avoir indiqué les biais potentiels générés par les méthodes quasi expérimentales, notamment dans la sélection des individus des groupes, pouvant être liés en partie à l'omission de variables déterminantes, elle situe l'évaluation randomisée comme la mieux à même de résoudre ces obstacles. Mobilisant les travaux de R. Lalonde, elle signale même la différence de résultats possibles entre

⁵³² Le Lasmass était le Laboratoire d'analyse Secondaire et de Méthodes Appliquées à la Sociologie, une unité mixte du CNRS en double tutelle EHESS et ENS, il est devenu maintenant le Centre Maurice-Halbwachs.

⁵³³ Marc GURGAND, *Les méthodes d'évaluation statistique des politiques publiques*, IRESO, Les Ateliers M2S3, 27 avril 2004.

⁵³⁴ « Avant-propos. Fondements, allocation et impact de l'aide », (2005) 19-2 *Revue d'économie du développement* 5- 6.

⁵³⁵ EVALUATION GAP WORKING GROUP, *When Will We Ever Learn? Improving Lives Through Impact Evaluation*, Center for Global Development, 2006.

⁵³⁶ DARES, *Colloque sur les expérimentations pour les politiques publiques de l'emploi et de la formation*, Paris, 22 mai 2008, p. 146.

⁵³⁷ Esther DUFLO, « Évaluer l'impact des programmes d'aide au développement: le rôle des évaluations par assignation aléatoire », (2005) 19-2 *Revue d'économie du développement* pp 185-226, p.188.

les deux méthodes⁵³⁸. Si des limites de cette méthode sont identifiées, comme les sélections de l'échantillon, l'attrition différentielle entre les deux groupes, les effets de débordements, ou les effets d'apprentissage et d'imitation, elle soutient que ces derniers sont aussi présents dans les autres méthodes quasi expérimentales.

Elle montre aussi les atouts politiques de cette méthode qui permet d'abord de mieux communiquer avec les hommes politiques. Elle estime que : « *les différences entre les bonnes et les mauvaises évaluations sans assignation aléatoires sont difficiles à communiquer, particulièrement aux décideurs politiques, à cause de toutes les mises en garde qui doivent accompagner leurs résultats. En pratique, ces mises en garde ne devraient jamais être données aux décideurs politiques, et même si elles sont données, elles devraient être ignorées ; dans les deux cas, les décideurs politiques ont des chances d'être induits en erreur. Cela suggère que, bien que les évaluations sans assignation aléatoire soient encore nécessaires, il devrait y avoir un engagement à mener des évaluations par assignation aléatoire lorsque c'est possible.* »⁵³⁹

D'ailleurs en mobilisant des exemples d'application au Mexique et au Kenya, elle soutient que « *les officiels du gouvernement ont reconnu la valeur de l'évaluation par assignation aléatoire et sont en fait préparés à payer pour cela* »⁵⁴⁰. Elle préconise cette méthode comme particulièrement appropriée pour l'action des Organisations Non Gouvernementales dont « *à la différence des gouvernements on n'attend pas qu'elles servent la population entière* », mais estime qu'elles nécessiteront l'« *assistance technique (par exemple de la part des universitaires) et d'un financement extérieur* ». Considérant qu'elles ne sont « *pas plus coûteuses que les autres types d'enquêtes et coûtent beaucoup moins cher que la poursuite de politiques inefficaces* », elle explique que leur développement peut être freiné si « *les partisans (d'un programme) bloquent les évaluations par assignations aléatoires puisqu'elles révéleraient les véritables impacts du programme aux votants* ».

Enfin, elle enjoint les agences internationales à promouvoir cette méthode, lui assignant le statut de « *bien public mondial* », dont le bénéfice de savoir si un programme fonctionne ou non s'étend au-delà du pays le mettant en œuvre⁵⁴¹. Elle propose que ces institutions mettent en place des « *incitations pour les évaluateurs* » et des « *fonds spécialisés pour encourager et financer des évaluations d'impact rigoureuses ou pour en diffuser les résultats* ». Elle suggère d'appliquer le modèle de la FDA (*Food and Drug Administration*), qui exige une évaluation par assignation aléatoire des effets d'un médicament avant d'être distribué, et de l'appliquer aux politiques de développement. Pour conclure, elle estime que cette nouvelle crédibilité établie par cette méthode pourra contrebalancer le scepticisme des bailleurs de fonds et construire un support de long terme pour le financement du développement.

Cette communication/publication est un exemple emblématique d'activité de problématisation contribuant à poser la méthode expérimentale comme une ressource aux crises rencontrées dans différents champs ou, pour reprendre le vocabulaire de M. Callon, comme un *point de passage obligé* pour la résolution des problèmes du champ.

⁵³⁸ *Ibid.* p. 200.

⁵³⁹ *Ibid.* p. 193.

⁵⁴⁰ *Ibid.* p. 203.

⁵⁴¹ Cette présentation prend d'autant plus de sens du fait que la « protection des biens communs de l'humanité » soit un des objectifs affichés de l'Agence Française de Développement, voir Jean David NAUDET et Jocelyne DELARUE, *Promouvoir les évaluations d'impact à l'Agence Française de Développement : renforcer l'appropriation et l'apprentissage institutionnels*, 02, coll. Série Notes méthodologiques Ex Post, Département de la Recherche, Division Evaluation et capitalisation, Agence Française de Développement, 2007, p. 5.

D'abord, E. Duflo restreint le débat de l'évaluation scientifique des impacts à celui de la démarche expérimentale ou quasi expérimentale de recomposition d'un groupe témoin. On note d'ailleurs, qu'elle parle de « *méthode d'évaluation par assignation aléatoire* », établissant comme évident son appartenance à la démarche expérimentale. Elle évince ainsi les démonstrations alternatives de la causalité - n'ayant pas recours à des scénarios contrefactuels - présentes dans le débat américain (voir chapitre 1). Cette méthode est présentée comme une solution politique à deux niveaux pour, d'une part, favoriser la conduite de programmes basés sur l'efficacité et dépenser moins en programmes inutiles et, d'autre part, légitimer l'intervention des agences internationales et des ONG - secteur en crise de légitimité - et pérenniser leur financement par des bailleurs de fonds.

La diffusion des travaux d'Esther Duflo, si elle ne connaît pas encore le succès des années suivantes, rencontre un fort succès auprès des économistes français. En témoigne la réception par cette dernière, en 2005, du prix du meilleur jeune économiste, alloué par le Cercle des économistes et le journal *le Monde*⁵⁴². Un statisticien explicite ce qui nous est apparu comme l'état d'esprit des chercheurs à cette période, « *je pensais que c'était infaisable en France. Je me disais quelle force institutionnelle pourra briser ? ...j'avais l'impression qu'il y avait des obstacles énormes pour le faire, ce qui était à la fois vrai, mais pas tout à fait car il fallait simplement que cette méthode... qu'on prouve qu'elle soit faisable* ». La recherche de la conduite d'une expérimentation aléatoire qui servirait d'exemple et ouvrirait la voie devient alors l'objet principal du travail institutionnel qui s'amorce. Dans cette quête, l'alliance politique apparaît peu à peu comme un élément indispensable à la composition d'un *champ d'action stratégique* stabilisé et efficace.

1.2. L'étincelle de l'émergence, composition d'un champ embryonnaire : rencontres, organisation et définition des intérêts

C'est dans le contexte de structuration des champs décrit ci-dessus que s'organisent les actions « d'agence » de ses promoteurs. Les actions scientifiques éparées autour de la promotion de cette méthode rencontrent en 2005 leur pendant politique. La pratique de « l'expérimentation sociale » apparaît comme une ressource commune offrant l'opportunité d'un nouveau positionnement dans les champs politiques et scientifiques respectifs de ces acteurs. L'émergence d'un *entrepreneur frontière* de taille, Martin Hirsch, marginal sécant inscrit à la fois dans le champ scientifique, politico-administratif et associatif, constitue l'élément catalyseur de ce *champ d'action stratégique*. Un travail institutionnel scientifique et politique s'organise autour de l'entreprenariat individuel et de la structuration de nouvelles organisations collectives dédiées à la promotion de cette méthode. Un travail de cadrage cognitif s'opère, ainsi qu'une mobilisation des ressources financières et un dépassement des barrières réglementaires. Le travail institutionnel de problématisation identifie précisément les crises du champ de l'économie appliquée et de l'action publique pour proposer l'expérimentation comme solution commune. Cependant, à ce stade, cette solution reste présentée de manière assez floue afin de mobiliser un maximum d'acteurs, en effet, elle ne précise pas encore de manière explicite son association avec la méthode aléatoire. La présentation de la « méthode expérimentale » fait des allusions suffisantes à la méthode aléatoire pour que les chercheurs se mobilisent en sa faveur, tout en ne figeant pas cette association pour que les acteurs de l'action publique, favorables à l'expérimentation

⁵⁴² Le Cercle des économistes est un groupement d'une trentaine d'universitaires ayant exercé dans le public ou le privé organisant des débats « conciliant réalisme politique, faits et rigueur des analyses ». Pour plus de précision voir leur site <http://www.lecercledeconomistes.asso.fr/qui-sommes-nous>

mais réfractaires au tirage au sort, l'acceptent. Le travail de ce champ d'action stratégique embryonnaire s'organise pour l'instant à l'écart des champs stratégiques de l'Administration centrale.

1.2.1. La commission « Famille, vulnérabilité et pauvreté », rencontre et amorce d'un projet commun

Dans son ouvrage autobiographique *Secret de fabrication*, Martin Hirsch raconte l'origine de cette Commission. Philippe Douste-Blazy, alors ministre des Solidarités, de la Santé et de la Protection sociale depuis mars 2004, pensait, à l'occasion d'un remaniement ministériel, prendre la direction du ministère du Budget, mais trouve ses compétences uniquement étendues à celle de la famille, en devenant en novembre 2004, Ministre des Solidarités, de la Santé et de la Famille. Souhaitant agir dans ce nouveau domaine de compétence, il sollicite Martin Hirsch pour qu'il lui soumette des propositions d'intervention, et ce dernier propose un travail par commission⁵⁴³. L'encadré ci-dessous revient sur le profil de ce nouvel acteur.

Encadré 6: Martin Hirsch, un entrepreneur aux multiples identités⁵⁴⁴

Martin Hirsch, un entrepreneur aux multiples identités

M. Hirsch est né le 6 décembre 1963 à Suresnes (Hauts-de-Seine).

Un héritage familial de science appliquée et d'engagement politique

- Son grand père Étienne Hirsch, grand résistant, directeur adjoint de l'armement pour les forces de la France libre, et proche de Jean Monnet, prend la direction du Commissariat au Plan en 1952 et le conserve jusqu'en 1959 où il préside la commission de l'Euratom.
- Par la suite, son père Bernard Hirsch, polytechnicien, est un des concepteurs de la ville nouvelle de Cergy-Pontoise dans les années soixante-dix, il prend la direction de l'école nationale des ponts et chaussées de 1983 à 1988.

Son histoire familiale est ainsi ancrée dans l'institutionnalisation de la science au service de l'action publique et de l'engagement.

Un cursus universitaire entre science et administration publique

Son parcours académique atypique le situe à l'intersection des champs de la médecine, de la recherche en biologie et de la haute fonction publique. Il a un engagement précoce, il milite au parti socialiste à 17 ans. Il cumule les formations ; pour reprendre ses propos, « à 20 ans, j'étais élève à l'ENS, en 3ème année de médecine et en licence et maîtrise de biochimie ». Il commence par des études de médecine (faculté de Cochin Port-Royal : 1981-1986) qu'il n'achève pas, et intègre en 1983 l'Ecole normale supérieure et obtient un DEA de neurobiologie et une maîtrise de biochimie en 1986. Il intègre l'École nationale d'administration (Ena) en 1988, jusqu'en 1990.

Une carrière professionnelle à multiples facettes

Durant sa carrière, il cultive ses différents réseaux et centres d'intérêts en alternant les postes à responsabilité dans la haute fonction publique, et dans les instances sanitaires.

Il sort de l'Ena en tant qu'Auditeur au Conseil d'État (1990) et assure en parallèle des cours à l'Institut d'études politiques de Paris (1990-1997). Il devient conseiller juridique à la Caisse nationale d'assurance maladie des travailleurs salariés (CNAMTS) de 1990 à 1992, puis conseiller juridique au ministère de la Santé (1992-1993) il est finalement promu Maître des requêtes au Conseil d'État en 1993) où il occupera la fonction de Secrétaire général adjoint du Conseil d'État jusqu'en 1995. Il devient professeur à l'Ena en 1994.

⁵⁴³ Martin HIRSCH, *Secrets de fabrication*, Paris, Grasset, 2010, p. 84.

⁵⁴⁴ Source : <http://www.ambafrance-cl.org/CV-de-M-Martin-Hirsch>; <http://www.letudiant.fr/loisirsvie-pratique/loisirsvie-pratique-people/martin-hirsch-15275.html>; http://www.sciencespo.fr/chaire-sante/sites/sciencespo.fr.chaire-sante/files/CV%20_HIRSCH_Martin.pdf

Après ce parcours dans la haute administration, il retourne à la santé publique en 1995 en devenant Directeur de la pharmacie centrale des hôpitaux à l'Assistance publique - hôpitaux de Paris (APHP)(1995-1997). Il devient la même année Président de l'Union centrale des communautés Emmaüs (1995-2002).

Il occupe ensuite un poste politique et directeur de Cabinet du secrétaire d'État à la Santé (B. Kouchner) et conseiller au cabinet de la ministre de l'Emploi et de la Solidarité (M. Aubry) (1997-1999).

Il retourne ensuite au milieu sanitaire et occupe le poste de Directeur général de l'Agence française de sécurité sanitaire des aliments (AFSSA) (1999-2005). Tout en occupant cette fonction, il assure la fonction de président d'Emmaüs France (2002-2007), Administrateur de l'Agence européenne de l'évaluation des médicaments (2002-2007), et de membre du forum de l'Autorité européenne de sécurité alimentaire et du Conseil national de lutte contre l'exclusion (CNLE) (2003-2007). Il devient aussi en 2004, membre du comité consultatif de la Haute Autorité de lutte contre les discriminations et pour l'égalité des chances.

La commission « Famille, Vulnérabilité et Pauvreté », organisée en 2005, marque le début de la mobilisation des acteurs autour de la promotion de la méthode expérimentale. Martin Hirsch, alors président d'Emmaüs France, préside les travaux de cette commission mise en place et réunie pendant 100 jours au premier trimestre 2005⁵⁴⁵. Celle-ci a pour but de réunir les différents acteurs des politiques en direction des familles défavorisées, afin de faire émerger des propositions de nouvelles modalités d'intervention.

Elle réunit des personnalités scientifiques, des représentants d'associations, du mouvement familial, des membres du conseil d'administration de la CNAF (Caisse Nationale des Allocations Familiales) et de la MSA (Mutualité Sociale Agricole). Elle est appuyée dans ses travaux par cinq directions centrales, 26 contributeurs extérieurs et bénéficie des « conseils et points de vue » de 17 personnalités⁵⁴⁶. Au total 93 personnes ont contribué aux travaux de cette commission ainsi qu'à la production du rapport restituant ces travaux, intitulé « *Au possible, nous sommes tenus. La nouvelle équation sociale, 15 résolutions pour combattre la pauvreté des enfants* ». On retrouve parmi ces contributeurs des personnes qui joueront un rôle central dans la promotion de la méthode expérimentale.

Des hauts fonctionnaires composent la colonne vertébrale du Cabinet de Martin Hirsch, comme Étienne Grass, alors à l'IGAS (Inspection Générale des Affaires Sociales), qui est un des trois rapporteurs de la commission, et Jean-Benoît Dujol, alors à la DSS (Direction de la Sécurité Sociale), une des directions centrales mobilisées en appui.

Des chercheurs déjà engagés, ou qui vont agir par la suite pour la promotion de la méthode d'évaluation expérimentale par assignation aléatoire : comme Éric Maurin (EHESS), lequel fait partie des personnalités qualifiées mobilisées, Yannick L'Horty (Université d'Evry) est contributeur, Marc Gurgand (CNRS) et Thomas Piketty (ENS), des personnalités dont on a sollicité les conseils et points de vue.

Selon les acteurs avec lesquels nous nous sommes entretenus, cette commission devient le lieu de rencontre d'individus qui vont sceller des alliances par la suite. Elle permet la

⁵⁴⁵ Martin HIRSCH, *Secrets de fabrication*, op. cit. p. 87.

⁵⁴⁶ Pour consulter la liste détaillée des participants à cette commission, se référer à la page 5 du rapport COMMISSION FAMILLES, VULNERABILITE, PAUVRETE, *Au possible, nous sommes tenus. La nouvelle équation sociale, 15 résolutions pour combattre la pauvreté des enfants*, 2005.

rencontre de Martin Hirsch avec Étienne Grass et Jean-Benoît Dujol⁵⁴⁷, et le début de leur discussion sur l'expérimentation sociale. Un de ces acteurs se souvient : « *on a fait un petit groupe à trois et c'est là qu'on a conçu, dans le cadre du rapport, le RSA, qu'on a imaginé de le tester et qu'on a beaucoup discuté des expérimentations dont on avait entendu parlé à l'époque, sachant qu'à l'époque il y avait une actualité universitaire académique, mais elle n'était pas beaucoup diffusée dans le monde politique. Nous, notre grande référence c'était l'expérimentation canadienne dans le self sufficiency project... l'idée qu'on puisse répliquer ce genre de choses en France me paraissait un peu bizarre, mais on trouvait ça génial* »

Bien que ce rapport soit principalement connu comme étant la première présentation des principes du RSA (Revenu de Solidarité Active), il est aussi le lieu de la première évocation de la démarche d'expérimentation sociale, alliant mode d'intervention politique et évaluation par une méthode scientifique. Ce rapport témoigne dans un premier temps, d'un nouveau discours sur l'action publique et définit de nouvelles modalités de l'intervention scientifique dans ce domaine. Il pose les bases de la légitimation du recours à l'expérimentation sociale que sont l'absence de mesure de l'efficacité de l'action publique, et la mobilisation d'experts non pour leur *savoir*, mais pour leur *savoir-faire*.

En introduction, les rédacteurs déplorent un état de crise politique, dû à « *l'inefficacité (étatique) par rapport aux dépenses* »⁵⁴⁸. S'incarne ici de manière particulièrement explicite la description d'une crise du champ stratégique de l'action publique et sa problématisation, positionnant la méthode expérimentale comme instrument susceptible de la résoudre que nous évoquons dans notre première hypothèse. Ce constat est un discours en plein essor dans les milieux de réflexion des économistes cette année-là, en témoigne la publication par le Cercle des économistes d'un Cahier consacré à la réforme de l'État⁵⁴⁹. Les contributeurs de ce travail se positionnent en porte à faux avec les travaux d'autres commissions : « *nous n'avons pas fait un rapport d'experts. Nous ne délivrons pas un savoir ou des solutions toutes faites* »⁵⁵⁰. En effet, ils soulignent le développement insuffisant de l'évaluation en France et malgré l'implication croissante des chercheurs, relèvent la difficulté de l'administration et des décideurs à accepter la conclusion de ces travaux. Se référant à l'exemple britannique de *l'evidence-based policy*, et du recours à l'expertise indépendante pour l'élaboration des politiques publiques, ils récusent la tendance française à internaliser ses instances d'expertise au sein de l'État⁵⁵¹.

Sur les 15 résolutions proposées pour combattre la pauvreté des enfants, une résolution entière est consacrée à l'expérimentation sociale, la Résolution n°13 intitulée : « *Audace, innovation, expérimentation: les mots clés des nouvelles interventions publiques* »⁵⁵². Cette résolution définit une crise du mode d'action traditionnel de l'État, en prenant à témoin son incapacité à résoudre la pauvreté latente

⁵⁴⁷ Selon l'entretien avec Jean Benoît Dujol. Néanmoins, Étienne Grass et Jean-Benoît Dujol étaient amis proches depuis Science-po.

⁵⁴⁸ COMMISSION FAMILLES, VULNERABILITE, PAUVRETE, *op. cit.* p. 6.

⁵⁴⁹ Patrick ARTUS (dir.), « Réformer l'Enseignement supérieur et la Recherche, un pacte pour une nouvelle université », *Le Cercle des économistes* 2004.5.135.

⁵⁵⁰ COMMISSION FAMILLES, VULNERABILITE, PAUVRETE, *op. cit.* p. 7.

⁵⁵¹ « *L'évaluation des politiques publiques est très insuffisante : des équipes de recherche s'impliquent de plus en plus, mais leurs conclusions, quand elles existent, demeurent mal admises par l'administration et les décideurs. (...) Au total, le contraste est particulièrement frappant entre le Royaume-Uni, où le gouvernement s'appuie sur une expertise indépendante pour construire ses politiques, et la France, où l'État se comporte comme s'il avait le monopole de l'expertise nécessaire à la décision sans se donner véritablement les moyens de procéder à une évaluation de son action, dont il tiendrait compte en temps utile.* » *Ibid.* p. 18.

⁵⁵² Cette résolution est développée de la page 96 à 100 dans le rapport, *Ibid.*

de ses citoyens⁵⁵³. Bien qu'elle reconnaisse le souhait de la France de s'engager dans l'expérimentation, la commission déplore son retard en la matière et l'absence de moyens consacrés⁵⁵⁴. Cette résolution constate l'incapacité de l'État à travailler de manière transversale, interministérielle, ainsi qu'à être assez réactif pour proposer des réformes sociales : *« un an pour élaborer une loi ; un an pour élaborer les décrets ; un an pour mettre en œuvre : c'est en général le temps pour déclarer qu'une réforme ne marche pas et en mettre en chantier une autre, sans même avoir correctement évalué les effets de la précédente »*⁵⁵⁵. Elle préconise alors le recours à de nouveaux modes d'action publique, davantage organisés autour de l'efficacité, et s'inscrivant dans la logique de la nouvelle Loi Organique sur Les Finances publiques. La seconde orientation de cette résolution propose de substituer une agence nationale de l'innovation et de l'expérimentation sociales à des organismes d'État, tels que les délégations interministérielles, estimant que l'État *« a du mal à accepter la logique expérimentale »*. Cette agence aurait pour but de *« mettre en place ces expérimentations et de suivre les projets soumis par des acteurs locaux »*⁵⁵⁶. Une organisation de l'expérimentation en dehors de l'État est donc envisagée. La troisième orientation de cette même résolution précise ce qui est entendu par le développement de *« l'expérimentation sociale »*. Après avoir identifié des domaines d'intervention à privilégier, comme notamment les modalités de l'organisation de l'accompagnement social, ce paragraphe propose deux manières de procéder, l'une descendante, l'autre montante :

- *« proposer des expérimentations selon des programmes qu'il (l'État) définirait lui-même, mais qu'il souhaiterait tester sur une partie du territoire ; dans ce cas, il lancerait une consultation, comme un appel d'offres, pour voir quelles sont les collectivités locales qui seraient prêtes à porter ensemble un projet expérimental et sous quelles conditions ;*
- *accepter les expérimentations dont l'initiative serait prise localement »*⁵⁵⁷.

Néanmoins, à aucun moment dans cette résolution n'est fait référence à la méthode scientifique d'évaluation expérimentale par assignation aléatoire. La quatrième orientation suggère de s'inspirer *« de la procédure irlandaise d'étude d'impact social »*, systématique en matière de pauvreté, et la sixième orientation propose de développer le *testing* social, mais la méthode aléatoire n'est pas mentionnée⁵⁵⁸. La seule référence à cette méthode est présente dans la proposition n°2 portant sur le RSA, laquelle évoque l'évaluation d'un dispositif similaire, le *Self Sufficiency Project* expérimenté en 1992 au Canada⁵⁵⁹. Cette première présentation de l'expérimentation sociale ne formule donc pas encore le lien entre ce nouveau mode d'action publique et la méthode scientifique d'évaluation aléatoire. Néanmoins, celle-ci est déjà présente dans les discussions et associée à l'expérimentation politique comme nous l'ont précisé les acteurs participants à cette commission⁵⁶⁰.

⁵⁵³ « La pauvreté est un miroir grossissant pour les personnes en difficulté. Elle l'est également pour mettre en évidence les limites de l'action traditionnelle de l'État. (...) Tout ceci conduit à préconiser de nouveaux modes d'action publique. » Ibid. p. 97.

⁵⁵⁴ « L'État prône l'innovation et l'expérimentation, mais il ne s'est pas donné les moyens de les mettre en œuvre. Notre pays est extrêmement pauvre en expérimentations. » Ibid.

⁵⁵⁵ Ibid. p. 97.

⁵⁵⁶ COMMISSION FAMILLES, VULNERABILITE, PAUVRETE, *op. cit.* p. 98.

⁵⁵⁷ Ibid. p. 99.

⁵⁵⁸ Ibid. p. 98-99.

⁵⁵⁹ La méthode n'est même pas nommée en tant que telle il est seulement mentionné que *« Les bénéficiaires du dispositif ont été sélectionnés au hasard »* COMMISSION FAMILLES, VULNERABILITE, PAUVRETE, *op. cit.* p. 34.

⁵⁶⁰ Cette question a été abordée directement lors de notre entretien avec M. Hirsch :

À la suite de ce travail, Martin Hirsch est sollicité pour entrer au gouvernement de D. De Villepin en juin 2005, mais il refuse cette offre. La proposition du RSA est peu à peu oubliée et Xavier Bertrand, ministre successeur de Philippe Douste-Blazy à la Santé et aux Solidarités, ne porte plus ce projet⁵⁶¹. Le président d'Emmaüs prend alors en charge lui-même le « *service après-vente* » de son rapport en constituant par son réseau le champ d'action stratégique de la méthode expérimentale⁵⁶².

1.2.2. Les instances d'une réorganisation de l'interaction entre science économique appliquée et action publique autour de l'expérimentation sociale

Suite au peu d'écho rencontré par les recommandations de ce rapport, Martin Hirsch prend en charge l'entrepreneuriat de ce champ d'action stratégique embryonnaire en mobilisant les membres des réseaux qu'il a tissés autant dans le champ scientifique que politique. Rapidement, la nécessité de la création de nouvelles organisations se fait sentir pour développer ce projet atypique, qui ne correspond pas aux normes de l'interaction de la science et du politique, comme elle s'était développée par le passé. Deux nouvelles entités hybrides, organisant la rencontre des acteurs du secteur privé, scientifique, et politique, sont créées suivant le modèle anglo-saxon. D'une part, l'Agence Nouvelle des Solidarités Actives organise la promotion et l'appui de la demande politique en expérimentation sociale. De l'autre, la création de *Paris School of Economics* - Ecole d'Économie de Paris - structure les capacités d'offres en évaluation de l'expérimentation sociale.

1.2.2.1. L'Ansa, le *travail institutionnel* politique hors de l'État

Martin Hirsch crée en janvier 2006 l'Agence Nouvelle des Solidarités Actives avec Benoît Guenuini, et une vingtaine de membres, certains ayant participé à la commission de 2005, dont Jean-Benoît Dujol, Étienne Grass et Éric Maurin⁵⁶³. Néanmoins, cette association n'est pas encore soutenue par l'État⁵⁶⁴. C'est pendant cette période que Martin Hirsch endosse son rôle d'entrepreneur frontière de la méthode expérimentale en France⁵⁶⁵. C'est une période d'activation intense de ses réseaux. Il s'engage dans une « *mobilisation tous azimuts* » et mobilise les anciennes équipes du ministère de la Santé

- (Chercheuse) : « dans le rapport il y a une recommandation claire qui est sur la conduite d'expérimentations et il n'y a pas encore de précision de méthode, est-ce qu'à ce moment-là vous aviez déjà envisagé la méthode expérimentale, l'application scientifique, ou c'était plus dans l'acte politique d'expérimenter ? »

- M. Hirsch : Non c'était les deux. Oui.

⁵⁶¹ C'est ici un euphémisme puisque Martin Hirsch rapporte dans son ouvrage les propos de ce nouveau ministre à ce sujet en ces termes « les RMIstes, pour qu'ils reprennent du travail, tu ne crois pas qu'il vaudrait mieux leur donner un bon coup de pied au cul, plutôt que de mettre de l'argent supplémentaire ? » Martin HIRSCH, *Secrets de fabrication*, op. cit. p. 89.

⁵⁶² Les propos exacts de Martin Hirsch sont « Nous faisons nous-mêmes le service après-vente de notre rapport » ASSOCIATION DES JOURNALISTES DE L'INFORMATION SOCIALE, « Martin HIRSCH - Président d'Emmaüs France et de l'Agence nouvelle des solidarités actives - Petit-déjeuner » (20 juin 2006).

⁵⁶³ Martin Hirsch a rencontré Benoît Guenuini, à Singapour début juillet alors qu'ils soutenaient tous deux la candidature de Paris aux JO de 2012. Les membres fondateurs de l'Ansa sont: Gaby Bonnand, Juliette Chevalier, Denis Clerc, Julien Damon, Jean-Benoît Dujol, Christophe Fourel, Benoît Guenuini, Étienne Grass, Nicolas Grivel, Paulette Guinchard, Martin Hirsch, Frédéric Martel, Éric Maurin, Serge Paugam, Pierre Rosanvallon, Fabien Tuleu, Sylvaine Villeneuve, Jean-Luc Volatier, Laurent Wauquiez. AGENCE NOUVELLE DES SOLIDARITES ACTIVES, « Revue succincte d'activité », 2006.1.

⁵⁶⁴ « Nous faisons comme si nous avions le droit et les moyens de le faire, mais nous nous heurtons aux réticences de l'ensemble des administrations, à l'exception de quelques francs-tireurs » Martin HIRSCH, *Secrets de fabrication*, op. cit. p. 90.

⁵⁶⁵ « Je me bats pour les expérimentations (...) Je trouve porte close » Martin HIRSCH, *Secrets de fabrication*, op. cit. p. 69.

qu'il a connu sous B. Kouchner, les membres de la commission Douste-Blazy ainsi que ses contacts dans le monde de l'entreprise⁵⁶⁶.

Il publie une tribune au Monde le 2 février 2006, dénonçant le manque d'intervention du gouvernement contre la pauvreté, et présentant l'expérimentation comme une nouvelle force d'intervention contre la pauvreté en mobilisant les exemples canadiens et britanniques⁵⁶⁷. Le premier ministre sollicite alors ses conseils, et M. Hirsch lui demande un portefeuille de 20 millions d'euros. Il n'obtient que deux millions d'euros, ainsi qu'une aide de l'administration, et « cinq lignes dans la loi » pour l'autoriser à amorcer les expérimentations. Deux amendements passent alors au vote parlementaire et lèvent les obstacles réglementaires à la conduite d'expérimentations :

- l'un, lors de la loi des finances de décembre 2006 (art. 142), autorise des dérogations aux textes en vigueur permettant aux départements d'expérimenter des dispositifs d'aide au retour à l'emploi, tels que le revenu de solidarité active, et prévoit aussi une procédure de candidature à l'expérimentation sociale pour des départements volontaires⁵⁶⁸.
- L'autre, en mars 2007, par l'article 52 de la loi instituant le droit au logement opposable, ouvre des possibilités d'expérimentations sur des contrats de travail aidés⁵⁶⁹.

Les départements volontaires à l'expérimentation doivent présenter « *avant le 30 juin 2007 un dossier décrivant les expérimentations envisagées, les objectifs poursuivis, les résultats attendus, les dispositions législatives et réglementaires auxquelles ils entendent déroger ainsi qu'un protocole d'évaluation* »⁵⁷⁰. L'obligation d'évaluation est établie pour les expérimentations ainsi que leur suivi par un comité d'évaluation. Chaque année les départements doivent adresser un rapport sur leur mise en œuvre au représentant de l'État dans le département contenant notamment des « *éléments relatifs à l'impact de ces mesures sur le retour à l'emploi* »⁵⁷¹. Néanmoins, la méthode pour établir l'impact n'est pas précisée.

L'Agence Nouvelle de Solidarité Active devient alors le catalyseur de l'émergence du champ d'action stratégique promouvant la méthode expérimentale. Au sein de l'Ansa s'organise et se théorise le couplage entre la dimension politique et scientifique de l'expérimentation sociale.

Lors de la première année de l'Ansa, un réseau de partenariat est constitué avec⁵⁷² :

- La DIISES (Délégation Interministérielle à l'Innovation à l'Expérimentation Sociale et à l'Économie Sociale), nouvellement créée en février 2006, qui offre un soutien financier

⁵⁶⁶ Marc LANDRE, « Les réseaux de Martin Hirsch », *L'Expansion.com* (1 mai 2006).

⁵⁶⁷ Martin HIRSCH, « Contre la misère, osons ! », *Le Monde* (2 février 2006).

⁵⁶⁸ « A titre expérimental et pour une durée de trois ans à compter de la date de publication du décret prévu au II du présent article, afin d'améliorer les conditions d'incitation financière au retour à l'emploi des bénéficiaires du revenu minimum d'insertion, l'État confie aux départements admis à participer à l'expérimentation la charge de financer la prime de retour à l'emploi instituée par l'article L. 322-12 du code du travail en tant que celle-ci est versée aux bénéficiaires du revenu minimum d'insertion ». *Loi n° 2006-1666 du 21 décembre 2006 de finances pour 2007*.

⁵⁶⁹ « II.-A titre expérimental, afin d'améliorer les conditions d'incitation financière au retour à l'emploi et de simplifier l'accès aux contrats de travail aidés, les départements mentionnés par le décret prévu à l'article L. O. 1113-2 du code général des collectivités territoriales sont autorisés, pour une durée de trois ans à compter de la date de publication dudit décret, à adopter, en faveur des bénéficiaires du revenu minimum d'insertion, tout ou partie des dérogations aux dispositions du code du travail et du code de l'action sociale et des familles prévues aux III et IV du présent article, dans les conditions fixées par les mêmes III et IV. » *Ibid.*, article 142.

⁵⁷⁰ *Ibid.*, article 142.

⁵⁷¹ *Ibid.*, article 142.

⁵⁷² AGENCE NOUVELLE DES SOLIDARITES ACTIVES, *Rapport d'activités, année 2006, 2007*.

pour 3 ans, établit un « *processus rigoureux* » de suivi des expérimentations, et documente les retours d'expériences et les évaluations des impacts ;

- La Caisse des Dépôts et Consignation, pour financer une partie des programmes dans le champ de l'innovation et de l'expérimentation sociale pour 3 ans ;
- L'assemblée des départements de France, pour sensibiliser les Conseils Généraux ;
- La République des idées, fondée par Pierre Rosanvallon en 2002, après la dissolution de la Fondation Saint Simon, laquelle est un « think tank » indépendant à la française, de gauche sociale démocrate, qui propose des analyses économiques et sociales de la société. Les idées développées dans ses publications sont reprises dans la campagne présidentielle à droite comme à gauche en 2007. Ce groupe recrute à l'époque un nombre important de chercheurs en laboratoires, lesquels formeront par la suite l'école d'économie de Paris, dont notamment Éric Maurin⁵⁷³. Ce partenariat aboutit à l'organisation du forum intitulé « la nouvelle critique sociale » à Grenoble en mai 2006 ;
- L'École d'Économie de Paris nouvellement formée, se lie à l'Ansa par une convention de 2 ans permettant « *la mise à disposition de chercheurs pour établissement de protocoles scientifiques de certaines expérimentations qui s'y prêteraient, afin d'évaluer de façon rigoureuse l'impact des expériences menées* ».

Encadré 7: La Délégation Interministérielle à l'Innovation, à l'Expérimentation Sociale et à l'Economie Sociale⁵⁷⁴.

La Diïeses

Créée en juillet 2000 auprès du ministre chargé de l'emploi et de la cohésion sociale, sous le nom de « délégation interministérielle à l'innovation sociale et à l'économie sociale », elle est devenue la « délégation interministérielle à l'innovation, à l'expérimentation sociale et à l'économie sociale » en février 2006. C'est cette direction interministérielle dont le manque d'activité en matière d'expérimentation avait été dénoncé par le rapport de la Commission Famille, vulnérabilité, pauvreté présidée par Martin Hirsch en 2005. On peut faire l'hypothèse que l'inclusion du terme « expérimentation sociale » dans son appellation n'est pas sans lien avec les recommandations et travaux de cette commission.

Elle a pour mission :

- de soutenir et de promouvoir au niveau national le développement des activités d'intérêt général innovantes pour l'élaboration et la mise en œuvre des politiques publiques ;
- d'étudier, de proposer et de coordonner, dans son champ de compétence, les mesures destinées à favoriser le développement des coopératives, des mutuelles, des associations et des organismes intervenant dans le même secteur et de promouvoir des expérimentations en ces domaines ;
- d'assurer une concertation entre les pouvoirs publics et les organismes intervenant dans le domaine de l'économie sociale ;
- de répartir, sous l'autorité du ministre chargé de l'emploi et de la cohésion sociale, les crédits du fonds d'innovation et d'expérimentation sociale gérés par les préfets. La délégation dispose de correspondants locaux.

Elle sera supprimée en janvier 2010, lors de la création de la Direction Générale de la Cohésion Sociale qui reprendra ses compétences notamment d'animation et de coordination des actions « en matière de politique d'innovation, d'expérimentation sociale et d'économie sociale et solidaire ».

Entre janvier 2006 et la nomination de Martin Hirsch au gouvernement en mai 2007, l'Ansa, via ses partenariats, organise la diffusion d'un discours autour de

⁵⁷³ Thierry LECLERE, « Ça s'appelle "La République des idées" », *Télérama* n°3047 (7 juin 2008).

⁵⁷⁴ Décret n°2006-151 du 13 février 2006 instituant une délégation interministérielle à l'innovation, à l'expérimentation sociale et à l'économie sociale., 2264 (2006), ORF n°38. Décret n° 2010-95 du 25 janvier 2010 relatif à l'administration centrale des ministères chargés des affaires sociales et portant création d'une direction générale de la cohésion sociale, (2010).

l'expérimentation sociale, intégrant en son sein l'expérimentation politique, l'expérimentation juridique et la méthode expérimentale par assignation aléatoire⁵⁷⁵.

Ce discours est exposé d'abord dans les numéros de RSA (Revue Succincte d'Activité), revue de l'Agence, dont le second numéro théorise ce qui est entendu par expérimentation sociale. Cette présentation reprend les critiques de l'action de l'État formulées par le rapport de la commission de 2005, en déplorant la lenteur de l'action publique et son manque de réactivité pour établir régulièrement qu'« *une mesure coûteuse a produit peu d'effets* »⁵⁷⁶. Ce numéro présente une « *démarche expérimentale* » afin d'« *aider à concevoir des politiques plus efficaces, en connaître l'impact réel, en évaluer les conditions de réussite, avant une éventuelle généralisation* »⁵⁷⁷. Dans un article intitulé « *L'expérimentation sociale : un nouveau mode d'action pour les politiques publiques* », Étienne Grass en présente les principes. Il réaffirme le lien de l'association avec l'École d'Économie de Paris nouvellement créée, laquelle a fait de « l'expérimentation sociale » un axe pédagogique. L'auteur se réfère aux démocraties anglo-saxonnes et relève le retard de la France, en faisant référence à la conduite de projets dès les années 60 dans les pays anglo-saxons, et à l'existence d'organismes d'évaluation tels que la *Rand corporation*, la *Manpower Demonstration Research Corporation* (MDRC) ou l'*Urban Institute* dont le travail d'inventaire note l'absence d'expérimentation en France, alors que plus de 17 pays y sont représentés⁵⁷⁸. L'effort de définition est explicite dans cet article, puisque l'auteur note l'approximation entourant le terme « d'expérimentation » et propose de revenir sur la « *réelle notion de l'expérimentation* »⁵⁷⁹. Se référant à la définition de l'*Urban Institute* il explique : « *une expérimentation sociale se définit par quatre caractéristiques : l'affectation aléatoire des bénéficiaires et d'un groupe de contrôle (random assignment), un dispositif de politique publique (policy intervention), un dispositif de suivi (follow up data collection) et une évaluation (evaluation)* ». Ces différentes composantes sont explicitées par la suite, dans une définition canonique. Une expérimentation sociale est⁵⁸⁰ :

1. « **Une innovation de politique sociale** ...(...) porteuse de sens pour les politiques sociales (...) qu'on en connaisse les présupposés et qu'on identifie les questions auxquelles elle est susceptible de répondre ». Le dispositif doit être « mûr » au moment de l'expérimentation et peu sophistiqué afin de pouvoir déterminer la part de « chacune de ses composantes dans les résultats obtenus » » ;
2. « **Initiée dans un premier temps à petite échelle compte tenu des incertitudes existantes sur ses effets...** : l'expérimentation est une mesure transitoire sur un petit groupe de population défini géographiquement ou sur la base d'autres critères ; elle doit aboutir à une décision politique; si les effets à attendre d'un nouveau dispositif sont relativement certains, cette dérogation au principe d'égalité (...) est peu justifiable » ;
3. « **Et mise en œuvre dans des conditions qui permettent d'en évaluer les effets...** : cela suppose que n'existent pas de biais dans la sélection des bénéficiaires et que ces bénéficiaires soient comparables à un groupe de non-bénéficiaires ; cela suppose

⁵⁷⁵ Jean-Claude BARBIER, « Evaluations "expérimentales": quelques leçons d'histoire », *Politiques sociales et familiales* 2012.110.19-31, p.20.

⁵⁷⁶ AGENCE NOUVELLE DES SOLIDARITES ACTIVES, « Revue succincte d'activité », 2006.2.

⁵⁷⁷ AGENCE NOUVELLE DES SOLIDARITES ACTIVES, « Revue succincte d'activité », 2006.2.

⁵⁷⁸ Il renvoie aux références suivantes: D. Greenberg, M. Schroder, *The digest of social experiments*, The Urban Institute press, 3ème ed, 2004; E. Duflo « Field Experiments in Development Economics », NBER, jan 2006;

D. Card, D. Hyslop, « Estimating the effects of a timelimited earning subsidy for welfare-leavers », SRDC working papers, fév 2005 ainsi qu'aux sites internet du <http://www.povertyactionlab.org/>; <http://www.mdrc.org/>; <http://www.urban.org/>; <http://www.campbellcollaboration.org/>

⁵⁷⁹ AGENCE NOUVELLE DES SOLIDARITES ACTIVES, « Revue succincte d'activité », 2006.2.

⁵⁸⁰ AGENCE NOUVELLE DES SOLIDARITES ACTIVES, « Revue succincte d'activité », 2006.2.

également que le groupe des bénéficiaires ait atteint une taille critique ; (...) cela suppose enfin qu'au moment où l'expérimentation est lancée son protocole d'évaluation soit arrêté et connu. »

4. « ...dans l'optique d'une généralisation », et donc veiller à la validité externe des résultats et la représentativité de l'échantillon ou les variations de contexte. Les travaux de recherche « récents » visant à établir des méthodologies pour « dépasser ou contrôler ces limites » sont évoqués, en prenant pour exemple ceux d'Esther Duflo.

Il mobilise des exemples étrangers vertueux, comme le programme *Sure Start*, en 2002 en Grande Bretagne, sur la prise en charge de la petite enfance dans les milieux défavorisés. L'auteur explique que toutes les questions juridiques posées par l'expérimentation sociale ne sont pas résolues. En effet, l'autorisation de conduite de dispositifs dérogatoires sur le territoire national renvoie aux règles « *classiques de dérogation au principe d'égalité (...). Cette jurisprudence permet des dérogations au principe d'égalité soit quand les situations des usagers sont elles-mêmes différentes, soit quand la dérogation ne « trouve sa justification que dans certaines nécessités d'intérêt général* » ». Selon Étienne Grass, on peut estimer que l'expérimentation correspond au second critère « *dans la mesure où la nécessité d'évaluer les politiques publiques de la façon la plus objective possible peut être reconnue comme une « nécessité d'intérêt général* » ». Quant aux droits relatifs aux bénéficiaires, il propose de s'inspirer des droits existants dans le domaine des essais cliniques, à savoir : le droit à l'information sur les modalités d'expérimentation, le principe du consentement éclairé avant d'entrer dans le dispositif, et le droit de sortir de l'expérimentation.

Le terme « d'expérimentations contrôlées » est employé, non pas pour parler explicitement de la méthode scientifique, mais de la conduite d'expérimentations en général. Cet article, écrit par Étienne Grass, fait acte de définition officielle de l'expérimentation sociale. Il a été ensuite repris à plusieurs reprises dans d'autres documents, parfois officiels, pour expliquer cette méthodologie.

Le travail de cadrage cognitif est aussi relayé pour le partenariat avec « *La république des idées* », et notamment l'organisation d'une conférence à Grenoble du 12 au 14 mai 2006 intitulé « *La nouvelle critique sociale* ». Cette manifestation était organisée en partenariat avec l'Ansa, la Région Rhône-Alpes, la ville de Grenoble, et le journal le Monde qui l'a précédée d'une série d'articles consacrés à la crise de la société française⁵⁸¹. Des documents précis de cet événement ne sont pas accessibles, mais des commentaires publiés offrent un aperçu de ce qui y a été discuté. Cette conférence, qui aurait réuni, selon ses organisateurs, plus de 8 000 personnes, conforte le diagnostic d'une société française en crise se heurtant à une faiblesse des analyses et, selon la République des idées, « *pour sortir de cette impasse, un double effort de connaissance et d'expériences s'impose* »⁵⁸². Elle plaide pour un passage à l'action en rassemblant « *à la fois les chercheurs, qui renouvellent la connaissance de la société, et les acteurs, qui renouvellent les pratiques* », en témoigne une présentation conjointe de l'état de la recherche sur la société française confrontée à des expériences associatives⁵⁸³. Le professeur de science politique Laurent Bouvet, commente cet événement sur son blog, il regrette que « *l'accent mis sur le travail essentiellement quantitatif de sociologues ou d'économistes a paru souvent reléguer au second plan des approches plus qualitatives ou plus réflexives sur la société* », ainsi que « *l'absence quasi générale de la politique*

⁵⁸¹ LA REPUBLIQUE DES IDEES, « La nouvelle critique sociale, Grenoble, du 12 au 14 Mai 2006 », *La République des idées*.

⁵⁸² *Ibid.*

⁵⁸³ LA REPUBLIQUE DES IDEES, *op. cit.*; Laurent BOUVET, « Forum "la nouvelle critique sociale" - Grenoble 12-14 mai 2006 », *L(B)LOG* (30 mai 2006).

dans les débats », absence aussi bien du personnel politique que de la prise en compte de l'action politique et de la logique des partis ou des syndicats⁵⁸⁴. Ce thème de dépolitisation de l'action publique est porté par la République des Idées, comme l'explique Thierry Pech dans le Monde du 20 mai 2006, laquelle association est citée par ce même auteur : « (...) nous voulons acquérir une lucidité panoramique sur la société d'aujourd'hui dans une totale indépendance vis-à-vis des partis politiques ».

En parallèle à ce travail cognitif, s'opèrent des changements institutionnels dans le milieu scientifique témoignant d'une recomposition des modes de financement de la recherche, et encourageant l'intervention de la science économique dans la sphère sociale et politique.

1.2.2.2. La fondation de l'Ecole d'Économie de Paris

L'entreprise de promotion de cette méthode par des acteurs scientifiques s'inscrit aussi dans une reconfiguration des organisations de recherche afin de répondre à une « crise » de la recherche française identifiée par ces derniers. Celle-ci est explicitée par un rapport du cercle des économistes de 2004. Dans ce rapport, ce groupe d'économistes décrit une « *crise du modèle universitaire* » français en posant notamment la question des moyens et de l'ouverture à des financements privés, mais surtout en appelant une restructuration de la recherche française. Tout en questionnant les critères du classement de l'Université Jiaotong, dit « de Shanghai », dont la première parution eut lieu la même année, ils émettent l'hypothèse que l'incapacité des universités françaises à dépasser le 100e rang tient à l'incapacité d'un système d'enseignement supérieur et de recherche à émettre des « *signaux lisibles ou perceptibles à l'étranger sur le marquage des formes d'excellence* », notamment dû à « *l'éparpillement* » des universités⁵⁸⁵.

Encadré 8: La genèse de l'école de l'économie de Paris, Paris School of Economics

La genèse de l'EEP-PSE

L'école d'économie de Paris est une entité aux fondations multiples. Elle est issue de la rencontre du laboratoire DELTA (fusion du Centre d'Économie Quantitative et Comparative de l'EHESS et du laboratoire d'Économie Politique de l'ENS), du laboratoire d'économie appliquée de l'INRA, du laboratoire des Sciences Sociales de l'ENS, puis du CEPREMAP et du CERAS sur le campus de Jourdan. En 2001 ces laboratoires se réunissent au sein de la Fédération PARIS-J (Pour l'Animation de la Recherche Inter-centre sur le Site Jourdan). S'en suit la création en 2004 d'une Unité Mixte de Recherche Paris-Jourdan, entre CERAS-DELTA- unités CNRS du CEPREMAP. Thomas Piketty est alors le directeur adjoint en 2005 de PARIS-J et devient celui de l'association préfigurant l'EEP. Il cherche du soutien auprès de Dominique de Villepin, alors Premier ministre, lequel souhaite créer l'équivalent français de la London School of Economics. Ce travail aboutit à la création des réseaux thématiques de recherche avancée (RTRA) et de la fondation de coopérations scientifique « PSE-Ecole d'économie de Paris » en décembre 2006. En outre des structures précitées, l'EEP réunit l'Université Paris-1, l'Ecole des ponts de Paris-Tech, la Région Ile-de-France et l'État, mais aussi des partenaires privés (AXA, Exane, l'AFPSE).



⁵⁸⁴ Laurent BOUVET, « Forum "la nouvelle critique sociale" - Grenoble 12-14 mai 2006 », *op. cit.*

⁵⁸⁵ Le classement de Shanghai est attribué sur la base de cinq critères équipondérés : le nombre de lauréats du Prix Nobel, le nombre de chercheurs fréquemment cités dans 21 domaines scientifiques, le nombre d'articles publiés dans les revues Nature et Science, le nombre d'articles mentionnés dans le « Science Citation Index » et dans le « Social Science Citation Index », l'effectif scientifique de l'établissement, Patrick ARTUS (dir.), « Réformer l'Enseignement supérieur et la Recherche, un pacte pour une nouvelle université », *Le Cercle des économistes* 2004.5.135, 59- 60.

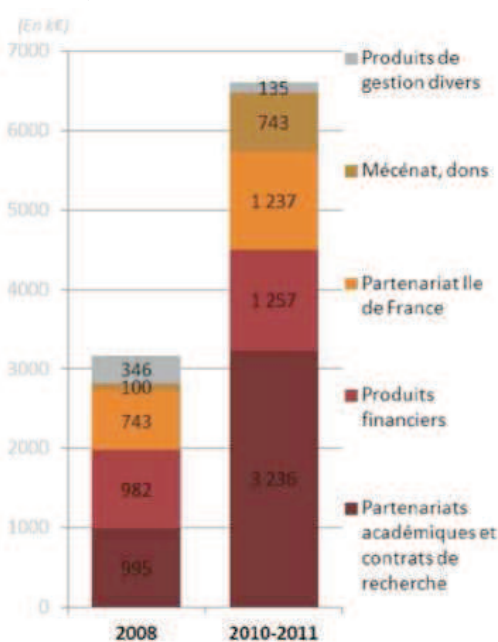
Les trois objectifs de la création de l'Ecole d'économie de Paris apparaissent comme une réponse à ce diagnostic. Ils sont rappelés par Roger Guesnerie, l'un de ses présidents⁵⁸⁶:

- Donner de la visibilité internationale par un regroupement de la « matière grise » ;
- « *Améliorer grâce à des financements supplémentaires, les conditions de recherche et rendre plus compétitive la place de Paris, en particulier plus attractive pour les chercheurs étrangers* ». En d'autres termes, avec son statut de fondation, cette école s'ouvre à des financements privés ;
- Tisser des liens durables entre l'économie et les autres sciences sociales.

Ayant célébré ses cinq ans en 2012, son ancien directeur estime que ces objectifs ont été atteints puisque l'École d'Économie de Paris est au 20^e rang mondial des départements d'économie, et au 5^e rang des non-Américains ; que le budget annuel de la fondation s'élève à sept millions d'Euros ; et qu'en 2012 elle réunissait 120 économistes, 300 étudiants en master, et 200 doctorants.

Ces partenariats privés et publics se manifestent par la constitution de Chaires thématiques financées par les institutions, comme la chaire Développement Durable (Ministère de l'Écologie, Développement durable, Transport et Logement), Macroéconomie internationale (Banque de France), et Économie de la Transparence et de l'Intégrité. Quatre groupes ont été développés depuis 2007, dont le G-MonD qui vise à approfondir l'analyse de la mondialisation et du développement, et rassemble à ce jour 4 partenaires-financeurs (AFD, Air France, Comité Colbert, Compagnie des Ecrehous)⁵⁸⁷ mais aussi par des collaborations ponctuelles. En cinq ans, une soixantaine de contrats d'expertise ont été réalisés à hauteur de 6,3 millions d'euros. Selon T. Piketty, « *à terme, les fonds privés dépasseront les financements publics* »⁵⁸⁸. Les enseignants, néanmoins, restent financés par leur corps d'origine.

Graphique 4: Évolution de la provenance des Fonds de PSE, en Kilo euros (2008-2011)



Sources : Paris school of economics, «Rapport d'activité 2010-2011, la science économique au service de la société», 2012, p 55

⁵⁸⁶ PARIS SCHOOL OF ECONOMICS, *La lettre: spéciale 5 ans*, n°8, janvier 2012.

⁵⁸⁷ François BOURGUIGNON, *PSE-Ecole d'économie de Paris : après 5 ans d'existence, un nouveau chapitre s'ouvre...*, Communiqué de presse, 16 février 2012. Elle reçoit aussi des financements d'une fondation de droit américain, l'American Foundation for the PSE, qui contribue pour un million d'euros. « Son président, Georges de Ménil, directeur d'études à l'Ehess depuis 1978 est héritier de la famille Schlumberger et fait partie de ceux qui ont quitté la France pour les États-Unis en 1981, au lendemain de l'élection de François Mitterrand. » Annie KAHN et Virginie MALINGRE, « Le Monde.fr : Les « French economists » font école », *Le Monde* (22 février 2007).

⁵⁸⁸ Annie KAHN et Virginie MALINGRE, « Le Monde.fr : Les « French economists » font école », *Le Monde* (22 février 2007).

L'école d'économie de Paris anime le LEEP (Laboratoire d'Économie Expérimental Parisien) qui est créé en 2007 au Centre d'économie de la Sorbonne et dont la principale activité est de mener des expériences d'économie en laboratoire⁵⁸⁹.

Néanmoins, en 2006, malgré le partenariat de cette fédération contracté avec l'Ansa, aucun Conseil général souhaitant conduire des expérimentations n'accepte d'avoir recours à l'évaluation expérimentale par assignation aléatoire. Les scientifiques souhaitant pratiquer cette méthode doivent alors s'organiser par eux-mêmes pour susciter l'intérêt politique, et obtenir son accord pour conduire les premières expérimentations.

1.2.2.3. La première expérimentation par assignation aléatoire à l'Agence Française de Développement

Le travail de promotion mené par Esther Duflo porte ses fruits dans le domaine du développement, en France, avec la conduite de la première expérimentation par assignation aléatoire française, mais sur sol étranger, par l'Agence Française de Développement (AFD)⁵⁹⁰. Dès 2005, L'AFD avait signé un accord avec le *Scientific Evaluation and Global Action* (SEGA), un centre de recherche de l'université de Californie (Berkeley et San Francisco) sur un projet de financement d'assurance santé au Cambodge (SKY), qui débute en 2008⁵⁹¹. Ce choix s'inscrit dans le mouvement international, décrit au chapitre 1, d'externalisation et de scientification de l'évaluation dans le domaine du développement. En témoigne l'intégration en 2006 de la division « évaluation » de l'AFD au département recherche de cette même institution⁵⁹². Néanmoins, ce n'est pas l'AFD directement qui décide d'avoir recours à la première commande de méthode d'évaluation expérimentale en France. C'est une institution de microcrédit au Maroc, Al Amana, qui a déjà sélectionné une équipe de recherche composée de membres du J-Pal, du MIT, et de l'École d'Économie de Paris (Bruno Crépon du Crest, Florencia Devoto, Esther Duflo et William Parienté), et qui sollicite l'AFD pour l'aider à financer cette évaluation⁵⁹³. L'évaluation de la microfinance est un réel enjeu dans le domaine de l'aide au développement puisque ce mode d'intervention a bénéficié d'un enthousiasme considérable sans être établi sur des faits probants, de plus, cette évaluation est une des premières à y appliquer la méthode expérimentale⁵⁹⁴. Des résultats rigoureux dans ce domaine constitueraient une ressource importante pour améliorer la position des chercheurs dans le champ de l'économie du développement.

Cette expérimentation débutée en 2006 marque le début de l'intérêt des institutions françaises pour cette méthode, mais ne s'inscrit pas néanmoins dans une pratique sur le territoire français. Or, la priorité pour les scientifiques souhaitant développer cette méthode est la démonstration sa faisabilité en France.

1.2.2.4. La première expérimentation aléatoire sur le territoire français

⁵⁸⁹ PARIS SCHOOL OF ECONOMICS, *Rapport d'activité 2010-2011, la science économique au service de la société*, 2012, p. 29.

⁵⁹⁰ L'AFD propose une assistance aux pouvoirs publics au secteur privé et réseaux associatifs en matière de développement. Ces trois objectifs sont la réduction de la pauvreté et des inégalités, la promotion de la croissance économique et la protection des biens communs de l'humanité.

⁵⁹¹ Jean David NAUDET et Jocelyne DELARUE, *Promouvoir les évaluations d'impact à l'Agence Française de Développement : renforcer l'appropriation et l'apprentissage institutionnels*, 02, coll. Série Notes méthodologiques Ex Post, Département de la Recherche, Division Evaluation et capitalisation, Agence Française de Développement, 2007, p. 14.

⁵⁹² *Ibid.* p. 7.

⁵⁹³ *Ibid.* p. 11.

⁵⁹⁴ Jean David NAUDET et Jocelyne DELARUE, *op. cit.* p. 11.

Au-delà des exemples étrangers d'incarnation de la méthode, les entrepreneurs français avaient besoin de mobiliser des exemples français de conduite de la méthode expérimentale par assignation aléatoire. Une poignée d'économistes du Crest lancent au début de l'année 2006 une initiative afin de pouvoir mettre en œuvre cette méthode sur le territoire⁵⁹⁵. Depuis le transfert des compétences d'accompagnement des RMistes aux départements en 2004, ces derniers peuvent ouvrir cet accompagnement à des opérateurs privés. Le département des Hauts-de-Seine décide en 2005 d'expérimenter cet accompagnement par des opérateurs privés et de l'évaluer rigoureusement et publie un appel d'offres⁵⁹⁶. Cette équipe d'économistes décide alors de répondre à l'appel d'offres d'évaluation en proposant la méthode expérimentale au moindre coût. Comme le souligne un des évaluateurs rencontrés, « *c'était vraiment très nouveau à ce moment-là, où on voulait maintenir le coût le plus faible possible puisqu'on voulait vraiment apporter la preuve que c'était possible de faire ça* »⁵⁹⁷. Ils proposent d'utiliser cette méthode sans la coûteuse ingénierie des enquêtes auprès des bénéficiaires, et appuient le suivi de cohortes sur les données de la Caisse des Allocations Familiales.

Au-delà de leur intérêt pour l'utilisation de cette méthode, le thème de l'expérimentation s'inscrit aussi au cœur d'un débat scientifique sur l'efficacité des dispositifs de conseil pour le retour à l'emploi des demandeurs d'emploi⁵⁹⁸. Alors que la majorité de la littérature scientifique conclut que le conseil des demandeurs d'emploi augmente l'accès à l'emploi dans plusieurs pays, seule une étude issue d'une expérimentation aléatoire aux Pays-Bas conclut à l'absence d'effet, et aucune n'a vraiment étudié le rapport coût-efficacité de ces prestations⁵⁹⁹. Cette expérimentation pourrait produire des résultats permettant de s'inscrire dans un débat scientifique international.

L'offre des scientifiques est retenue. Rapidement, des difficultés de mise en œuvre surviennent, le commanditaire n'ayant pas *a priori* pesé les exigences de rigueur de cette méthode. Nous ne reviendrons pas ici sur les détails de cette mise en œuvre, mais cette dernière, notamment à cause d'un très faible taux de participation, a abouti à une faible précision statistique endommageant la rigueur et la force des résultats produits⁶⁰⁰. La tendance des résultats est de conclure à un effet global nul du dispositif, mais d'une forte efficacité lorsque ciblé sur certaines populations, notamment celles les plus éloignées de l'emploi⁶⁰¹. Cette expérimentation ne s'est donc pas avérée pertinente comme exemple de bonne pratique, pour nourrir l'argumentaire et illustrer la conduite d'une expérimentation dans les règles de l'art, d'autant plus que les commanditaires n'ont pas autorisé les évaluateurs à mentionner leur participation à cette expérimentation afin d'éviter une récupération médiatique dans le contexte de la participation du président de ce Conseil général à la campagne présidentielle⁶⁰².

⁵⁹⁵ Cette évaluation a été conduite par Bruno Crépon (Crest-Insee et Jpal), Marc Gurgand (École d'économie de Paris, Crest Insee et Jpal), Thierry Kamionka (Crest-Insee) et Laurent Lequien (Dares et Crest-Insee)

⁵⁹⁶ « Accompagnement renforcé des bénéficiaires du RMI, dans les Hauts de Seine, France », *Abdul Latif Jameel Poverty Action Lab*.

⁵⁹⁷ Cet extrait d'entretien est corroboré par une communication de B. Crépon lors d'un colloque de la Dares « Par exemple, pour le RMI, on avait un système d'information très pauvre, parce que l'on voulait absolument se lancer dans cette opération et avoir un tout petit budget » DARES, *Colloque sur les expérimentations pour les politiques publiques de l'emploi et de la formation*, Paris, 22 mai 2008, p. 130.

⁵⁹⁸ Bruno CRÉPON, Marc GURGAND, Thierry KAMIONKA et Laurent LEQUIEN, *Is Counseling Welfare Recipients Cost-Effective? Lessons from a Random Experiment*, p. 2.

⁵⁹⁹ *Ibid.*

⁶⁰⁰ Cette expérimentation et les modalités de sa mise en œuvre font l'objet d'une étude détaillée en troisième partie, pour voir les détails de ces résultats lire *Ibid.*

⁶⁰¹ DARES, *Colloque sur les expérimentations pour les politiques publiques de l'emploi et de la formation*, op. cit. p. 127.

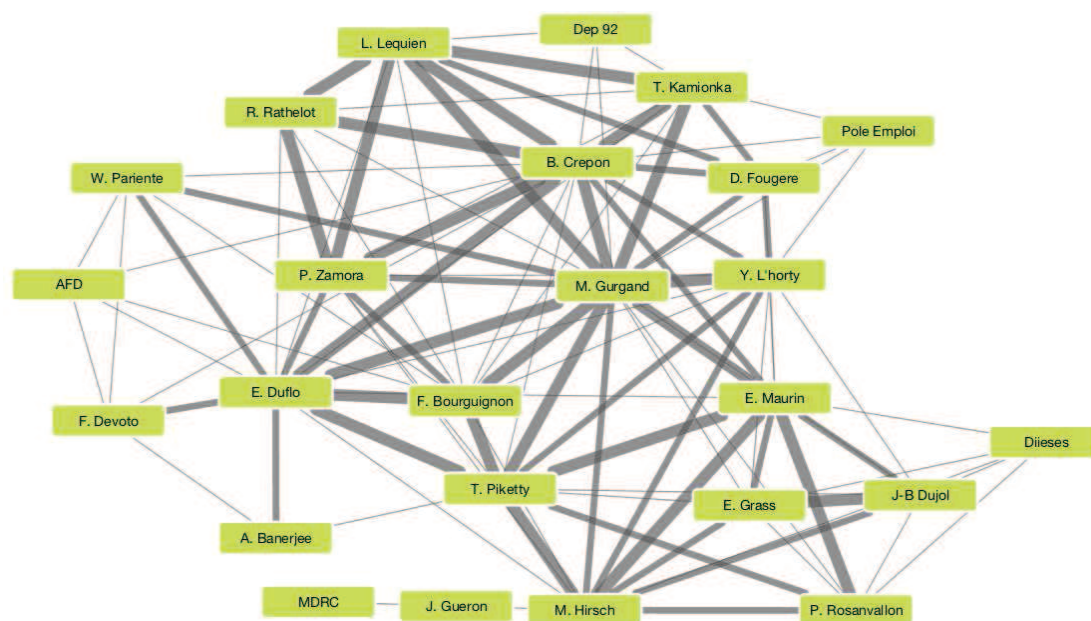
⁶⁰² Source : entretiens avec les évaluateurs de cette évaluation. Cette expérimentation est en effet une évaluation fantôme qui apparaît sur le site du J-pal Europe, mais auquel aucun acteur se réfère, préférant souvent citer les expérimentations de Pole Emploi comme les premières incarnations de la méthode expérimentale par assignation aléatoire en France.

On peut ainsi voir se dessiner à la fin de l'année 2006, l'émergence d'un *champ d'action stratégique* autour de la promotion de l'expérimentation sociale. Cet ensemble d'acteurs, composé de scientifiques, d'intellectuels, de hauts-fonctionnaires et de quelques hommes politiques, gravitant autour de La République des Idées, et de nouvelles organisations comme l'Ansa et l'École d'Économie de Paris ainsi que des structures plus anciennes comme le Crest, effectuent le *travail institutionnel* nécessaire pour l'avancement de cette méthode. Les obstacles réglementaires et financiers sont levés, et le travail de cadrage cognitif peut être observé. Ce dernier, connaissant quelques variantes, véhicule une conception d'une action de l'État français en crise accompagnée par un recours à l'expertise jugé non pertinent. Il promeut une solution passant par la rencontre des scientifiques et de l'action publique autour notamment de l'expérimentation sociale, dont la définition, formulée par ce champ, réunit méthode politique et méthode scientifique de l'évaluation expérimentale par assignation aléatoire. Cette dernière constitue une ressource de taille pour renforcer les positions des acteurs scientifiques engagés dans cette entreprise.

1.2.2.5. Cartographie des acteurs du Champ d'action stratégique émergent

Comme nous l'avons souligné dans notre partie méthodologique, nous avons eu recours à l'analyse de réseau afin de dessiner la cartographie des acteurs du champ, leurs relations et leurs positions. Nous n'avons pas étudié toutes les dimensions permises par cet outil d'analyse ce qui aurait pu constituer le travail d'une thèse à part entière. Afin d'identifier les membres de ce réseau nous avons croisé nos données issues d'entretiens et de l'analyse documentaire pour repérer les individus qui pendant cette période, de 2005 à 2007 (avant l'accession de M. Hirsch au gouvernement) ont contribué à la promotion de la méthode expérimentale, soit en la pratiquant, soit en publiant à ce sujet ou en présentant des communications ou participant à des commissions s'y rapportant. Nous avons identifié à cette période une trentaine de personnes ou d'organisations ayant contribué à ce travail institutionnel. Les liens entre ces individus sont de quatre types : participation à une même communication, scientifique ou institutionnelle, participation à une même expérimentation, publication en commun scientifique ou institutionnelle, et appartenance à la même organisation. Evidemment, ces liens sont en partie incomplets puisque nous n'avons pas eu systématiquement accès aux liens informels entre ces acteurs. Néanmoins, cette analyse nous donne un aperçu de la nature des liens de ce champ. La représentation ci-dessous nous montre l'épaisseur des liens entre chaque individus correspondant au nombre de fois où nous avons put recenser un des liens énoncés ci-dessus.

Figure 7 : Représentation réticulaire du champ d'action stratégique émergent de l'expérimentation par assignation aléatoire de 2005 à 2007 (voir document agrandi en annexe 3)

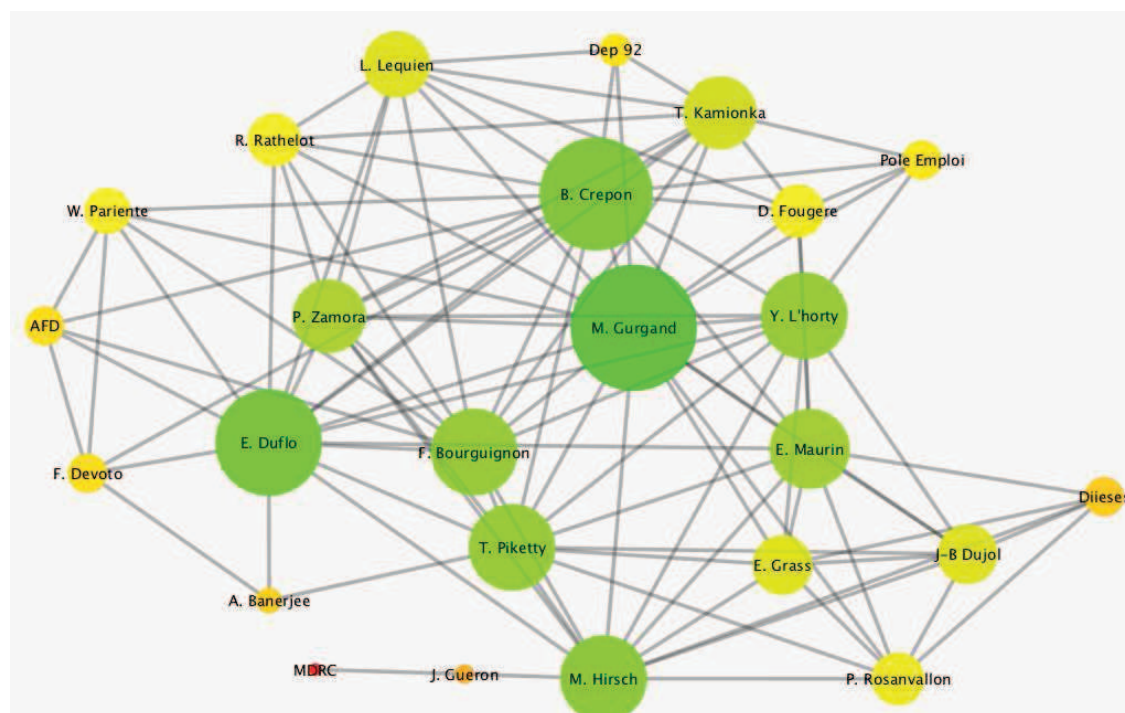


Pendant cette période on observe un champ assez restreint, composé de peu d'individus mais assez étalé. On remarque un ensemble d'individus déjà très connectés vers le haut du diagramme qui correspond au cercle des économistes statisticiens appliqués dans le domaine de l'emploi, soit du laboratoire du Crest de l'INSEE et de la DARES⁶⁰³. Au bas du diagramme apparaît un autre pôle constitué d'acteurs politiques ou des chercheurs investissent dans un rôle de conseiller du pouvoir. Au milieu de ces deux pôles, se trouvent les acteurs liant entre eux ces deux groupes d'acteurs, qui sont notamment les chercheurs de l'Ecole d'Economie de Paris comme M. Gurgand ou E. Maurin ou F. Bourguignon.

Nous avons choisi de centrer notre analyse sur deux mesures. D'abord le nombre de liens direct qu'un acteur détient avec d'autres membres du réseau qui est représenté par la taille du cercle, plus le cercle est grand plus la personne est connectée à un nombre important d'autres individus. Ensuite nous avons mobilisé une mesure appelée « *average shortest path length* » (longueur moyenne du plus court chemin) calculant la proximité de chaque acteur avec l'ensemble des membres du réseau. Cette mesure consiste à regarder pour chaque individu du réseau le plus court chemin pour aller vers tous les autres individus et à en faire la moyenne. Plus le score est faible plus l'individu est bien connecté au reste du réseau. Cela nous permet de faire émerger les *titulaires* de ce nouveau champ émergent qui se trouve au centre de la circulation d'information. Sur le diagramme ci-dessous, plus le nœud représentant l'individu est vert plus son score est faible, et donc plus il est connecté directement aux autres membres du réseau, l'inverse s'il est rouge.

⁶⁰³ Il est néanmoins important de souligner que le type de liens pris en compte peut participer aussi à biaiser l'analyse en faveur de liens plus importants entre les chercheurs puisque leur lien est rendu explicite par des publications en commun par exemple. Les liens entre les acteurs politiques sont beaucoup plus difficiles à identifier et quantifier mis à part leur participation aux mêmes commissions ou aux mêmes organisations.

Figure 8: Représentation réticulaire avec calcul de « *average shortest path length* » du champ d'action stratégique émergent de l'expérimentation par assignation aléatoire de 2005 à 2007 (voir document agrandi en annexe 3),



On observe que pendant cette période les individus les plus connectés à l'ensemble du réseau – caractérisés par un point vert – sont principalement des économistes comme B. Crépon du Crest, E. Duflo du MIT, M. Gurgand (PSE) ou Y. L'horty du TEPP. Martin Hirsch se présente comme le seul acteur de l'action publique étant bien connecté à l'ensemble du réseau. Les autres organisations gouvernementales, comme la Diieses, ou l' AFD sont beaucoup plus isolées. Les individus très reliés au reste du réseau sont ici les entrepreneurs principaux de ce nouveau champ d'action stratégique émergent. Nous allons élaborer la même analyse sur la période 2007-2009, ce qui nous permet de constater l'évolution de l'ampleur et de la nature du réseau et d'apprécier l'évolution des places de chaque individu.

2. Tisser une toile d'alliances autour d'un nouveau champ d'action stratégique : l'épreuve du pouvoir (2007-2009)

Suite à la période d'émergence décrite ci-dessus, s'ouvre un travail institutionnel d'une autre nature. L'accession de Martin Hirsch, l'entrepreneur frontière, à un poste au gouvernement donne une nouvelle ampleur à ce travail de promotion, en rendant accessibles des ressources financières, normatives et administratives, ainsi qu'offrant une tribune sans précédent au travail cognitif. S'engage alors, un travail institutionnel des titulaires du champ pour relier ce dernier à d'autres champs pouvant être source de stabilisation, et élargir son influence. Ce procédé s'inscrit dans un processus d'institutionnalisation de ce champ par la sécurisation de ses ressources. En outre, la défense de la méthode expérimentale comme instrument de *evidence-based policy* correspond à un exercice de légitimation d'un positionnement stratégique délicat dans le champ politique, d'un acteur aux valeurs de gauche dans un gouvernement à majorité de droite. Ce recours au pragmatisme permet notamment d'importer les règles légitimatrices du champ scientifique à un acteur refusant de s'inscrire dans les règles d'affrontement du champ politique partisan. Or, la priorité pour le champ d'action

stratégique est l'incarnation « d'alliés » sur le territoire français, c'est-à-dire des exemples de la conduite de la méthode expérimentale par assignation aléatoire, afin d'en démontrer l'applicabilité et la pertinence. La pratique de l'expérimentation aléatoire est ainsi à la fois la fin et le moyen de l'émergence de ce champ d'action stratégique. Le développement suivant revient sur ces initiatives qui contribuent à renforcer l'interdépendance entre les acteurs scientifiques et politico-administratifs promouvant cette méthode. Ils rencontrent aussi un travail institutionnel déstabilisant, inscrit dans les pratiques et les règles du champ de la politique traditionnelle. À la fin de cette période se dessinent les contours clairs de ce champ d'action stratégique, dont les principaux acteurs sont identifiés et dont le discours est clairement et publiquement diffusé. Nous étudions dans un premier temps, les initiatives des membres « politiques » du champ d'action stratégique qui via le cabinet de Martin Hirsch s'engagent dans des activités d'acculturation de l'administration à l'expérimentation sociale, notamment par la conduite du dispositif du RSA (Revenu de Solidarité Active) et d'appels à projets. Dans un deuxième temps, nous mettons à jour les initiatives scientifiques au sein des champs de l'expertise d'Etat, et particulièrement au sein de la DARES, où ils bénéficient d'un accès privilégié. Ce travail d'acculturation est prolongé à l'extérieur de l'Etat par l'organisation d'événements de sensibilisation des milieux scientifiques et de l'action publique ou sociale.

2.1. La méthode expérimentale, raison d'être, ressource et objectif de l'accession au pouvoir politique

Ce développement revient sur les initiatives politiques pour développer la méthode expérimentale par assignation aléatoire. Cette méthode est au cœur de l'activité politique de l'accession au pouvoir du pendant politico-administratif de *ce champ d'action stratégique* émergent. Cette « mission », qui légitime leur action gouvernementale, passe dans un premier temps par des obligations de compromis méthodologiques fragilisant les liens avec les acteurs scientifiques de ce champ. Suite à la prise en compte des obstacles cognitifs à la réalisation de leur mission, les acteurs de ce *champ d'action stratégique* s'engagent dans un travail institutionnel cognitif d'une nouvelle ampleur et mobilisent de nouveaux moyens d'intéressements pour garantir l'incarnation française de cette méthode et inscrire ce nouveau champ dans des relations de proximité avec d'autres champs à même de sécuriser ses ressources. Après l'occasion manquée de l'expérimentation du RSA, la conduite d'appel à projets apparaît comme l'outil à privilégier. Il préfigure déjà une reconfiguration des rapports traditionnels entre le champ scientifique et politique.

2.1.1. Le cabinet politique comme laboratoire, l'ancrage du champ d'action stratégique dans des champs étatiques

La période de 2005 à 2007, évoquée précédemment, a pour particularité de s'inscrire dans un contexte de pré-campagne et de campagne politique pour les élections présidentielles françaises de mai 2007. Martin Hirsch, en tant qu'entrepreneur frontière qui gagne à garder ouvertes ses affiliations, s'applique à ne soutenir aucun des candidats tout en promouvant le RSA auprès des deux partis majoritaires pour assurer son succès⁶⁰⁴. Le RSA fait partie du programme de campagne autant de la candidate du Parti

⁶⁰⁴ Martin HIRSCH, *Secrets de fabrication*, op. cit. p. 26.

socialiste que du candidat de l'UMP (Union pour un Mouvement Populaire)⁶⁰⁵. Après son accession à la présidence de la République, Nicolas Sarkozy propose à Martin Hirsch de rentrer au gouvernement. Dans un premier temps, celui-ci refuse et propose en échange de fournir la documentation pour mettre en place le RSA sans son concours. Martin Hirsch relate cette sollicitation, qui se fait de plus en plus pressante, avec la perspective d'un levier pour le RSA :

(Nicolas Sarkozy) « Si vous entrez au gouvernement, votre RSA, il est voté cet automne et mis en place au 1^{er} janvier »

(Martin Hirsch) « Justement, je ne pense pas que cela soit raisonnable. Le RSA, il faut l'expérimenter avant de le généraliser, sans brusquer les acteurs de terrain et dans des conditions qui permettent de vérifier et de mesurer son efficacité. Pour cela, il faut dix-huit mois. »⁶⁰⁶

La démarche expérimentale, ainsi qu'un budget conséquent pour le RSA, font ainsi partie des conditions de l'entrée de Martin Hirsch au gouvernement, nommé Haut Commissaire, par le décret du 19 mai 2007, du Haut Commissariat aux Solidarités Actives⁶⁰⁷. Personnalité d'« ouverture » dans un gouvernement à droite de l'échiquier politique, le Haut Commissaire et son cabinet doivent se forger une posture légitimatrice de leur action au sein de ce gouvernement et l'expérimentation est leur argument principal pour justifier d'une action non partisane basée sur les résultats.

Martin Hirsch accepte l'entrée au gouvernement en négociant un statut particulier. Il est entré sous un « *mini contrat de gouvernement* », circonscrit au RSA, et obtient un Haut Commissariat l'attachant au Premier ministre et à aucun autre ministère⁶⁰⁸. D'ailleurs, la dénomination de sa fonction n'est pas anodine, puisqu'elle définit davantage la « *position d'un haut fonctionnaire ou d'une autorité administrative que d'un responsable politique* »⁶⁰⁹. Le décret du 12 juin délimite précisément ses fonctions⁶¹⁰. La lettre de mission du Président l'enjoint à « réconcilier nos compatriotes avec l'action politique *en prouvant* qu'elle peut encore changer les choses » et pose comme objectif pour le gouvernement de « réduire d'au moins un tiers en cinq ans la pauvreté dans notre pays » il souhaite que « le revenu de solidarité active soit créé à la fin de l'année 2008 et si possible avant. ». L'expérimentation est mentionnée ainsi que son évaluation qui appuiera l'élaboration d'un projet de loi⁶¹¹. L'éducation est un second axe d'intervention, puisque Martin Hirsch est aussi mandaté d'un travail avec le ministre de l'Éducation nationale pour accompagner la scolarité « des enfants pauvres ». La démarche expérimentale est revendiquée explicitement par la présidence comme s'inscrivant dans la démarche de la Révision Générale des Politiques Publiques de l'été 2007 (RGPP)⁶¹². La RGPP a pour objectif d'évaluer les résultats des interventions

⁶⁰⁵ Le RSA est proposé explicitement dans « Pacte présidentiel », le programme du parti socialiste, proposition 19 « *Mettre en place un revenu de solidarité active (RSA)* », il est évoqué explicitement dans le programme de Nicolas Sarkozy par la formulation « *Je ferai en sorte que les revenus du travail soient toujours supérieurs aux aides sociales* »

⁶⁰⁶ Emmanuel LEVY, « Miracle chez Hirsch: la Résurrection de Saint Martin », *Marianne*.

⁶⁰⁷ Décret du 18 mai 2007 relatif à la composition du Gouvernement, 9714 (2007), JORF n°115. Martin HIRSCH, *Secrets de fabrication*, Paris, Grasset, 2010, p. 33.

⁶⁰⁸ Martin HIRSCH, *Secrets de fabrication*, op. cit. p. 37.

⁶⁰⁹ Martin HIRSCH, « Martin Hirsch : « Je ne suis ni dupe ni naïf » », *Le monde* (20 mai 2007), p. 6.

⁶¹⁰ Avec l'appui de plusieurs ministères, le décret précise qu'il « prépare la réforme des minima sociaux » « il prépare la réforme des contrats aidés et des mécanismes d'incitation à la reprise d'activité et en suit la mise en œuvre », il mène « des programmes de lutte contre la pauvreté » « il prépare les travaux du comité interministériel de lutte contre les exclusions et du Conseil national des politiques de lutte contre la pauvreté et l'exclusion sociale et il organise les travaux du Conseil national de l'insertion par l'activité économique ». Décret n° 2007-1008 du 12 juin 2007 relatif aux attributions déléguées au haut-commissaire aux solidarités actives contre la pauvreté, 10273 (2007), JORF n°135.

⁶¹¹ Accent dans la citation rajouté par l'auteur, *Lettre de mission de M. Nicolas SARKOZY, Président de la République, adressée à M. Martin Hirsch, Haut commissaire auprès du Premier ministre, aux Solidarités actives contre la pauvreté*, 2007.

⁶¹² « Nous comptons sur vous pour que soient conduites, à votre initiative, les démarches innovantes, parfois expérimentales, susceptibles de renouveler l'action publique sur ce point. (...) Réussir les réformes attendues par les Français et cesser la spirale

publiques afin d'améliorer la qualité du service en allouant les moyens aux politiques les plus utiles⁶¹³. La grammaire de *l'evidence-based policy* est donc explicitement présente sans que ce terme ne soit mentionné pour autant. On retrouve dans le ton de cette lettre de mission le même discours présenté précédemment diagnostiquant une crise de l'action de l'État et l'appel à un nouveau mode d'intervention pragmatique, basé sur l'efficacité, et permettant d'arbitrer entre les interventions efficaces et inefficaces dans un objectif de contrôle des dépenses de l'État.

Cette entrée au gouvernement de Martin Hirsch est fortement critiquée et entraîne la scission de plusieurs alliances du champ d'action stratégique élaboré précédemment. De nombreuses personnalités de gauche le vivent comme une trahison. Pierre Rosanvallon et, de fait, la République des idées, lui retire son soutien. Le Haut Commissaire légitime sa position par un travail ciblé sur une mission, ainsi qu'une neutralité partisane. Il souligne dans son ouvrage dédié à sa période au gouvernement, « *j'ai vite compris que je n'étais fait ni pour l'engagement partisan ni pour la discipline de parti. J'ai toujours eu envie de contribuer à des politiques publiques, mais je ne me suis jamais senti attiré par la politique elle-même* »⁶¹⁴. L'efficacité de son action devient alors sa seule source de légitimité envers la majorité de droite et l'opposition de gauche.

Il s'entoure d'une quinzaine de personnalités « *neutres* » issues de la commission Famille, Vulnérabilité, Pauvreté, des personnes « *compétentes, dévouées à la cause, désintéressées et passionnées, soudées et solidaires* », avec, selon le Haut Commissaire, pour ossature le trio Emmanuelle Wargon, Étienne Grass et Jean-Benoît Dujol⁶¹⁵. Les quatre mots d'ordre donnés à cette équipe sont d'être loyaux et fiables, irréprochables sur le plan technique, de toujours avoir une idée d'avance et d'être craints. Selon Martin Hirsch, la « technique expérimentale » était leur principal outil pour avoir cette longueur d'avance⁶¹⁶. Les membres de ce cabinet défendent cette manière de conduire la politique publique. Jean-Benoît Dujol, par exemple, défend dans la presse l'intérêt de cette méthode scientifique s'intéressant à l'impact de la structure des incitations sur les comportements des individus, en revenant notamment sur les exemples de son utilisation au Canada⁶¹⁷.

Les membres du cabinet se positionnent en technocrates à la charnière de la recherche et de la politique. Ce cabinet ne s'inscrit pas dans les champs traditionnels de l'État constitués de ministères délimités. Cette équipe est isolée au milieu de l'action gouvernementale « *Ils disaient (les conseillers ministériels) : "Ces gars, c'est une secte." C'est revenu aux oreilles de Martin. Il a adoré !* »⁶¹⁸. Engagés sur une mission, leur démarche est atypique pour un cabinet ministériel. Des observateurs soulignent que, contrairement à d'autres cabinets, leur feuille de route était précise et orientée vers un même objectif⁶¹⁹. Cet objectif est double, puisqu'il tient à la fois à la mise en place du RSA, mais avant tout à la conduite de l'expérimentation de ce dernier, ainsi que d'autres expérimentations dans leur domaine de compétences. Ils doivent alors faire face à la

de l'endettement ne sont nullement inconciliables, mais sont au contraire deux objectifs complémentaires dès lors qu'il est décidé d'abandonner les politiques qui ne marchent pas au profit de politiques qui marchent. » *Lettre de mission de M. Nicolas SARKOZY, Président de la République, adressée à M. Martin Hirsch, Haut commissaire auprès du Premier ministre, aux Solidarités actives contre la pauvreté*, 2007.

⁶¹³ *Ibid.*.

⁶¹⁴ Martin HIRSCH, *Secrets de fabrication*, op. cit. p. 16.

⁶¹⁵ *Ibid.* p. 38. voir membres du cabinet en annexe

⁶¹⁶ *Ibid.* Martin p. 40.

⁶¹⁷ Jean-Benoît DUJOL, « Incentives matter », *Le tigre* 6 (octobre 2007), p. 14-15.

⁶¹⁸ Propos de Jean-Benoît Dujol recueillis dans la presse « Jean-Benoît Dujol, Un jeune au service des jeunes », *Les Echos* (3 juin 2010), p. 11.

⁶¹⁹ Un observateur explique en entretien « *le cabinet se distinguait des autres car il savait où il allait. et l'administration centrale se disait : « eux au moins, on sait où ils vont », je pense que ça a été assez fort* »

difficulté d'agir en dehors des champs traditionnels de l'État. L'expérimentation sociale devient alors une ressource stratégique pour occuper une place privilégiée dans l'action du gouvernement.

2.1.2. Le RSA, rendez-vous manqué de l'expérimentation, le politique contre la méthode scientifique

La première mission de ce cabinet est la conduite de l'expérimentation du RSA. Celle-ci a pour vocation de tester la mise en œuvre de cette réforme, mais aussi d'éprouver sur un dispositif de grande ampleur la méthode expérimentale par assignation aléatoire. La logique expérimentale s'inscrit ici, dans une expérimentation de modalités de mise en œuvre et non dans une logique de pilote dont la généralisation est conditionnée par les résultats de l'expérimentation, étant donné que la « généralisation » du RSA est déjà prévue dans la lettre de mission du président, comme précisé plus haut. Néanmoins, cette expérimentation, au final non aléatoire, ne peut constituer un exemple de bonne pratique d'expérimentation à mobiliser pour nourrir le travail institutionnel cognitif du champ d'action stratégique émergent, puisqu'elle n'est ni une incarnation de l'utilisation de cette méthode scientifique, ni un exemple dans les règles de l'art de décision politique éclairée par un résultat scientifique.

L'expérimentation du RSA regroupe les deux types d'expérimentations prévus par la loi. C'est d'une part une expérimentation de transfert de nouvelles compétences (la gestion de la prime pour l'emploi) qu'on envisage de pérenniser ultérieurement aux Conseils Généraux, et d'autre part une expérimentation de dérogation puisque les départements sont habilités à déroger à certaines dispositions - en changeant la périodicité du versement de la prime - pour tester plusieurs modalités de mises en œuvre⁶²⁰. Cette démarche conforte encore le mouvement commun liant la décentralisation des compétences, l'expérimentation et l'évaluation.

Un calendrier prévisionnel très serré répondant aux impératifs du temps politique est fixé⁶²¹. Le comité d'évaluation est rapidement mis en place. Il est composé de représentants de départements, des services de l'État et de « *personnalités qualifiées dont la compétence est reconnue en matière d'évaluation des politiques publiques* »⁶²². Ses fonctions sont : d'appuyer les départements dans la conduite des évaluations locales, de constituer un cadre comparatif d'ensemble pour les expérimentations locales pour permettre une évaluation nationale, et de mettre en place des dispositifs d'évaluation externe pour compléter les données issues des fichiers de gestion, et enfin de rédiger les rapports d'évaluation qui seront soumis au Parlement.

Sur les dix personnes qualifiées le composant, la moitié sont des chercheurs engagés dans la promotion de la méthode expérimentale par assignation aléatoire et composant le champ stratégique embryonnaire actif avant l'arrivée au pouvoir de Martin Hirsch. En témoigne notamment la présence des chercheurs de l'Ecole d'économie de Paris, comme

⁶²⁰ Gwénaële CALVES, *Histoire et perspectives du cadre juridique des expérimentations législatives et réglementaires en France*, Paris, Dares, Colloque sur les expérimentations pour les politiques publiques de l'emploi et de la formation, 22 mai 2008, p. 95- 106, p. 100.

⁶²¹ 20 juin 2007 : présentation du RSA au Conseil des ministres ; 10 juillet 2007 : installation du Comité d'évaluation national des expérimentations du RSA ; du 10 au 13 juillet 2007 : discussion du projet de loi à l'Assemblée nationale ; 30 septembre 2007 : date limite d'inscription pour les départements volontaires à l'expérimentation du RSA ; Expérimentation par les départements de la vague 1 en novembre 2007, expérimentations par les départements de la vague 2 début 2008, fin 2008 : généralisation du dispositif à l'ensemble du pays CNFPT, *Les dossiers documentaires des Conférences, Vers un revenu de solidarité active ? Conférence du 5 décembre 2007*, CNFPT première couronne, 5 décembre 2007.

⁶²² Ce comité s'est réuni 5 fois environ chaque semestre entre juillet 2007 et juillet 2008

Éric Maurin, Marc Gurgand, et François Bourguignon qui occupe la présidence ; un chercheur du Crest, Pierre Cahuc ; et un du Centre d'Étude et de l'Emploi, Yannick L'Horty⁶²³. Le choix des membres de ce comité affiche donc explicitement la volonté d'expérimenter le RSA par la méthode aléatoire contrôlée. Ce choix est explicité par un des membres du Cabinet de M. Hirsch : *« Quand il [M. Hirsch] a été nommé au gouvernement et qu'il a créé ce cabinet, lui était déjà assez au clair sur le type d'expérimentation qu'il voulait mettre en œuvre, en particulier la question de la méthode aléatoire était d'une certaine manière déjà tranchée dans son esprit il restait à convaincre les interlocuteurs qui allaient être impliqués [...] « comme on s'orientait vers des méthodes expérimentales basées sur la randomisation on a voulu avoir des experts sur ces questions là ».*

La visibilité nationale de l'expérimentation du RSA est un moyen de diffuser la construction cognitive élaborée précédemment autour de l'expérimentation. L'expérimentation est alors redéfinie en recouvrant par le même terme la démarche d'évaluation et le procédé juridique. Un glissement s'opère en permanence entre le mode d'intervention politique et la méthode d'évaluation scientifique. Le président du comité d'évaluation estime que l'expérimentation du RSA permet d'introduire *« la pratique de l'évaluation par expérimentation »*, la présentant comme *« basée sur la comparaison d'un échantillon aléatoire de zones géographiques ou d'individus exposés à la réforme étudiée et d'un échantillon de zones ou d'individus dans le statu quo, elle constitue une méthode pratiquement irréfutable d'évaluation »*⁶²⁴. Néanmoins, il est surprenant qu'ici le RSA soit présenté comme une application de cette méthodologie scientifique à grande échelle alors que, comme le souligne Yannick L'Horty par ailleurs, *« cette expérimentation ne correspond pas à ce que les spécialistes appellent une expérience « randomisée » (avec affectation au hasard dans un groupe test et un groupe témoin). Ni la liste des départements expérimentateurs, ni le périmètre des zones tests dans chaque département, ni la liste des bénéficiaires du RSA n'ont été choisis au hasard. »*⁶²⁵

En effet, la volonté du président du comité d'évaluation et du Haut Commissaire d'avoir recours à la méthode expérimentale par assignation aléatoire pour évaluer l'expérimentation du RSA s'est heurtée aux présidents de Conseil Généraux. Un observateur interne raconte cette rencontre entre deux cultures : *« Concrètement les chercheurs sont arrivés en disant : « bon c'est simple on va découper votre département en 6-7 zones, on va voir et on tirera au sort une zone sur deux, et on mettra en place dans les zones et dans d'autre non ». Alors les autres (présidents des Conseils Généraux) ont ouvert de grands yeux en disant : « non, non, nous on met en place là où on a envie de le mettre, et vous vous démerdez pour l'évaluer ». Et donc il a fallu trouver un juste milieu, c'est plutôt les présidents de CG qui ont gagné, beaucoup de chercheurs sont partis à ce moment, certains sont restés en se disant que ça valait quand même la peine de faire quelque chose même si ça n'était pas dans l'épure. ».*

Devant l'impératif politique, les scientifiques ont dû faire marche arrière et mobiliser une méthode quasi expérimentale de sélection des territoires comparables, élaborée par Yanick L'Horty. Le RSA ne peut donc pas constituer cet exemple de grande ampleur,

⁶²³ Les membres de ce comité sont François Bourguignon (Ecole d'Economie de Paris), Pierre CAHUC (CREST), Marc GURGAND, (EEP), Gilbert LAGUANELLE, Secours Catholique, Yannick L'HORTY, Centre d'Étude et de l'Emploi, Claude MARTIN, CNRS- ENSP, Elizabeth MAUREL, UPMF de Grenoble, Éric MAURIN, EHESS, Renée THOMINOT, comité des usagers d'Angers, Jérôme VIGNON, Commission Européenne, COMITE D'EVALUATION DES EXPERIMENTATIONS, *Rapport d'étape sur l'évaluation des expérimentations RSA*, Paris, 2008, p. 3.

⁶²⁴ François BOURGUIGNON, « Le RSA montre la voie de l'évaluation », *Les Echos* (3 juin 2009).

⁶²⁵ Yannick L'Horty, « Paradoxe de l'évaluation du RSA », *Projet 1*, n° 308, C.E.R.A.S (2009): 26.

indispensable pour démontrer que cette méthodologie peut être appliquée en France et qu'elle permet de produire des résultats rigoureux indiscutables par les idéologies politiques. De plus, la démarche explicitée de l'expérimentation sociale, qui laisserait le temps à une évaluation rigoureuse pour décider de la généralisation du programme, n'est pas respectée.

L'expérimentation est prise dans les pratiques du champ politique et dans un calendrier de programmation des réformes ouvrant une fenêtre de temps très limitée pour faire passer des réformes de cette ampleur. Un membre de Cabinet du Haut Commissariat explique : « *On avait un problème de calendrier, c'est qu'il a fallu très vite alimenter la réflexion sur l'application de la loi pour des raisons de calendrier politique. Il y avait un créneau pour faire passer une loi d'envergure de ce type qui ne se représenterait pas forcément un an ou deux plus tard ; ce qui d'un point de vue scientifique aurait été le temps idéal pour pouvoir vraiment tirer tous les enseignements des expérimentations* ». Les premiers résultats intermédiaires sont donc produits rapidement pour septembre 2008, afin d'éclairer une discussion au parlement 15 jours après. Très rapidement, les acteurs politiques comme Xavier Bertrand et Martin Hirsch se saisissent de ces résultats intermédiaires et déclarent dans une tribune du journal le Monde que « *le RSA a fait ses preuves sur le terrain* ». Cet argument sera repris par Martin Hirsch lors de la présentation du projet de loi devant l'Assemblée nationale où il déclare : « *Nous nous appuyons sur des résultats, sur des faits, sur des témoignages, sur des trajectoires, (...) et non sur une idéologie. (...) Le comité d'évaluation a estimé que cet écart était statistiquement significatif* »⁶²⁶. On retrouve ici clairement une mobilisation des ressources appartenant normalement au champ scientifique pour étayer cette démonstration et augmenter la légitimité politique.

Le RSA s'appuie en effet sur cette légitimité scientifique. Comme le notent B. Gomel et E. Serverin, le titre même de la loi « *généralisant le revenu de solidarité active et réformant les politiques d'insertion* » se réfère à la fois à l'opération de généralisation scientifique d'étendre des conclusions au-delà du champ de l'expérimentation, et l'opération institutionnelle qui applique une dérogation ouverte à des collectivités territoriales à l'ensemble du territoire⁶²⁷. D'ailleurs on trouve dans la loi une référence explicite à cette légitimité scientifique : « *Près d'une année d'expérimentation apporte des informations positives, avec des taux d'emploi supérieurs de 30 % en moyenne à ceux que l'on constate dans les zones témoins. Ces données qui sont désormais consolidées et franchissent les seuils de significativité statistique sont tout à fait encourageantes. Elles sont présentées dans le rapport du comité d'évaluation annexé à ce projet de loi* »⁶²⁸.

La relation avec les chercheurs investis dans le comité d'évaluation du RSA se trouve fragilisée par cette utilisation de leur travail. Le président du comité scientifique, bien qu'il comprenne les nécessités politiques de généralisation, estime qu'elle a été décidée trop tôt et que de nombreux points devaient être éclaircis, puisque les évaluations intermédiaires « *étaient entachées de biais importants et que la probabilité d'erreur était trop élevée pour en tirer des enseignements* »⁶²⁹. Yannick l'Horty renchérit dans un

⁶²⁶ Pour un retour détaillé sur cet argumentaire développé lors de cette deuxième séance du jeudi 25 septembre 2008 se référer au papier de Bernard GOMEL et Evelyne SERVERIN, « Expérimenter pour décider? Le RSA en débat », *Document de travail Centre d'étude de l'Emploi* 2009.119.

⁶²⁷ *Ibid.* p. 14.

⁶²⁸ *LOI n° 2008-1249 du 1er décembre 2008 généralisant le revenu de solidarité active et réformant les politiques d'insertion.*

⁶²⁹ Cette position est résumée dans deux articles des Echos « (le journaliste) Le rapport intermédiaire du comité, l'an dernier, évoquait un taux de retour à l'emploi plus élevé de 30 %...François Bourguignon : J'avais dit à l'époque que les évaluations intermédiaires étaient entachées de biais importants et que la probabilité d'erreur était trop élevée pour en tirer des enseignements. » « les conditions de cette généralisation ont été décidées bien trop tôt dans le processus

article: « *il est légitime de s'interroger sur l'utilité d'un dispositif d'évaluation aussi ambitieux. La généralisation du RSA a en effet été décidée avant que toutes les conclusions de la phase d'expérimentation ne soient tirées. Les résultats de l'étude d'impact n'étaient pas disponibles lorsqu'il a été décidé de généraliser le dispositif* »⁶³⁰. Concluant par : « *Les efforts déployés par les évaluateurs pour produire des résultats robustes ne sauraient dédouaner le politique de sa responsabilité.* »⁶³¹ Cette position est d'ailleurs renforcée lors de la production du rapport final en mai 2009 concluant à une revue à la baisse de l'estimation de l'impact des expérimentations invalidant les résultats intermédiaires⁶³².

Au final le président du comité d'évaluation estime avoir « *essuyé les plâtres* » de l'expérimentation⁶³³. Mais l'expérimentation du RSA entre dans une logique de diffusion du cadre cognitif de l'expérimentation aléatoire, d'acclimatation, comme l'explique Martin Hirsch plus tard dans une communication où il revient sur cette volonté de procéder par tirage au sort : « *Les éléments éthiques, voire juridiques, viennent s'y opposer. Alors, nous avons essayé de les contourner, tout en sachant que de ce fait on affaiblissait la force de nos protocoles mais que l'on pouvait acclimater les acteurs à cette logique.* »⁶³⁴.

2.1.3. L'appel à projets APDIESES : le difficile contrôle de l'intéressement par le travail cognitif

Parallèlement à cette initiative, le cabinet du Haut Commissaire tente de « *pousser des acteurs à se lancer dans des programmes expérimentaux* »⁶³⁵ et annonce, dès septembre 2007, le lancement d'un appel à « projets innovants » avec le soutien de la DIIIESES⁶³⁶. Cet appel à projets a pour objectif de « *faire émerger les idées, protocoles, expérimentations, susceptibles :*

- *De répondre à des questions concrètes posées par l'organisation des dispositifs sociaux relativement à l'efficacité des politiques sociales, aux raisons d'échecs ou de succès de certaines politiques ;*
- *De proposer des dispositifs innovants dans le champ de l'égalité des chances, de l'éducation, des modes de garde, de réduction des inégalités de santé... A ce titre, les expérimentations proposées pourront avoir trait à des dispositifs de traitement (dispositif*

d'expérimentation » Etienne LEFEBVRE, François BOURGUIGNON et Martin HIRSCH, « La crise rend la réforme plus indispensable que jamais », *Les Echos* (22 mai 2009); François BOURGUIGNON, « Le RSA montre la voie de l'évaluation », *Les Echos* (3 juin 2009).

⁶³⁰ Yannick L'HORTY, « Paradoxe de l'évaluation du RSA », (2009) 1-308 *Projet 21* - 29, p. 28.

⁶³¹ *Ibid.* p. 29.

⁶³² « On ne peut donc pas dire que l'application du RSA garantit un taux de retour à l'emploi plus élevé de 9 % partout et tout le temps. Mais, statistiquement, la probabilité que l'on se trompe quand on dit que l'effet emploi est positif est faible (12 %). Cela laisse peu de place à l'incertitude. », COMITE D'EVALUATION DES EXPERIMENTATIONS, *Rapport final sur l'évaluation des expérimentations RSA*, Paris, 2009.

⁶³³ « Bien sûr, nous avons essuyé les plâtres. Nous aurions souhaité disposer de plus de temps pour évaluer l'impact des expérimentations, travailler sur davantage de zones, pouvoir les choisir de manière plus aléatoire. » François BOURGUIGNON, « Le RSA montre la voie de l'évaluation », *Les Echos* (3 juin 2009).

⁶³⁴ DARES, *Colloque sur les expérimentations pour les politiques publiques de l'emploi et de la formation*, op. cit. p. 7.

⁶³⁵ C'est en ces termes que Martin Hirsch décrit cette initiative lors du colloque de la Dares organisé en mai 2008 DARES, *Colloque sur les expérimentations pour les politiques publiques de l'emploi et de la formation*, op. cit. p. 7.

⁶³⁶ Martin HIRSCH, « *Au possible nous sommes tenus* », dossier de presse, Haut Commissaire aux solidarités actives, 20 septembre 2007, p. 17. La DIIIESES avait déjà organisé deux appels à projets plus confidentiels l'un en 2006 auprès de 5 Régions et limité aux dispositifs d'aide à la personne et l'autre début 2007. L'appel à projet de 2007 avait un budget de 6 millions d'euros la sélection des projets avait lieu le 15 mars, les projets innovants étaient mis en œuvre entre avril et août et une valorisation et évaluation (sans précision de méthode) était prévue en septembre octobre 2007, PREFECTURE DE REGION RHONE ALPES, *Appel à pistes d'innovation et à expérimentation sociales 2007, financé sur le Fonds d'innovation et d'Expérimentation Sociales de la DIIIESES*, 2007.

ciblé sur les individus), d'incitation (actions destinées à modifier le comportement des individus, parfois à l'aide d'incitation financière), ou d'information (dispositifs modifiant le niveau d'information des individus). »⁶³⁷

Cet appel à projet s'adresse aux « équipes de recherches, les collectivités, les administrations, les entreprises, les associations »⁶³⁸. Il est abondé par le fonds d'innovation et d'expérimentation sociale (FIES) créé en 2006 qui a reçu une nouvelle dotation de 8 millions d'euros en 2008 pour le programme « Lutte contre la pauvreté : RSA et expérimentations sociales », pour accompagner le développement d'expérimentations en amont des politiques publiques en faveur de l'insertion et de la solidarité. L'appel à projets est doté d'un peu plus de 5 millions d'euros. Le texte de l'appel à projets précise qu'il s'adresse aux « projets pour lesquels le recours à la technique expérimentale est utilisé », et il indique ce que le Commissariat entend par méthode expérimentale en annexe⁶³⁹.

Il vise trois types d'expérimentations :

1. « Tester l'évolution de dispositifs de politiques publiques ou de pratiques privées(...) ;
2. Tester et évaluer un nouveau mode de fonctionnement pour un dispositif déjà existant(...) inférieur à 100 000 € ;
3. Les évaluations bénéficiant du soutien du haut commissaire : elles correspondent à des situations où la réponse à l'appel d'offre vise principalement à permettre l'accès à des données d'évaluation ou à débloquer une situation administrative, concernant un dispositif innovant. Pour ce type de projets, une aide à la valorisation du projet, notamment dans le cadre de publications scientifiques, sera prévue. ».

Une variété de domaines d'interventions est ouverte : « Réduction des inégalités de santé (en santé physique et mentale) ; Insertion professionnelle des jeunes 18 à 25 ans ; L'accompagnement professionnel et le développement des formations à destination des personnes éloignées de l'emploi ; Le développement de l'accueil du jeune enfant pour les familles en situation de pauvreté ; La promotion de l'égalité des chances dans l'éducation. »

Les porteurs de projet doivent soumettre une manifestation d'intention avant le 9 novembre 2007⁶⁴⁰. Si des critères d'appréciation ainsi que des indicateurs d'évaluation doivent être identifiés, il n'est pas requis de se présenter avec un évaluateur, puisque le Haut Commissariat propose une mise en contact avant le dépôt de candidature le 1^{er} janvier 2008⁶⁴¹.

Les exigences scientifiques de l'expérimentation sont précisées en annexe. Il est d'ailleurs expliqué que le comité de sélection peut faire reposer ses avis sur une appréciation par un « pair » pouvant être un chercheur étranger ayant une pratique de l'expérimentation sociale. Le document annexe présentant l'expérimentation sociale reprend quasiment mot pour mot l'article d'Etienne Grass dans la Revue Succincte

⁶³⁷ Martin HIRSCH, « Au possible nous sommes tenus », dossier de presse, Haut Commissaire aux solidarités actives, 20 septembre 2007, p. 17.

⁶³⁸ Ibid.

⁶³⁹ LE HAUT COMMISSAIRE AUX SOLIDARITES ACTIVES CONTRE LA PAUVRETE, *Appel à projet d'expérimentation sociale 2008*, 2007.

⁶⁴⁰ Il est précisé qu'un dossier de candidature pourra être déposé sans manifestation d'intention. Néanmoins il ne sera accepté que s'il met en relation un porteur de projet et une équipe d'évaluation

⁶⁴¹ LE HAUT COMMISSAIRE AUX SOLIDARITES ACTIVES CONTRE LA PAUVRETE, *Appel à projet d'expérimentation sociale 2008*, 2007, p. 3.

d'Activité publiée par l'Ansa en décembre 2006⁶⁴². Une différence minime, mais notoire pour notre étude, apparaît dans le paragraphe sur l'évaluation et précise qu'il est nécessaire qu'il « *n'existe pas de biais dans la sélection des bénéficiaires et que ces bénéficiaires soient comparables à un groupe de non-bénéficiaires* » ; une note de bas de page est rajoutée précisant qu'« *un débat existe sur la nécessité de procéder à une affectation aléatoire de ces bénéficiaires pour que ces critères soient respectés* » en faisant référence notamment aux écrits de J. Heckman « *Randomization and Social Policy Evaluation* »⁶⁴³.

Les porteurs de projet rencontrés soulignent néanmoins que cet appel à projets manquait d'informations pour comprendre les exigences des critères de sélection, « *on avait ni le vade-mecum, ni les outils, c'était extrêmement artisanal, on ne pouvait pas téléphoner pour savoir ce qu'il en était, c'était très sommaire au niveau des attendus de l'expérimentation sociale* ». Le haut commissariat organise, en soutien de cet appel à projet, une conférence de lancement du Grenelle de l'insertion du vendredi 23 et samedi 24 novembre 2007 afin d'explicitier ce qu'il entend par démarche expérimentale. Cette conférence témoigne de la consolidation du *champ d'action stratégique* autour du couplage de l'expérimentation et de la méthode expérimentale par assignation aléatoire. « *Les rencontres de l'expérimentation* » sont organisées par le Haut commissaire aux solidarités actives contre la pauvreté, la ville de Grenoble, et l'Ansa. On observe ici, une pérennité des partenariats créés en 2006, à l'exception de la République des Idées qui n'a pas soutenu la participation au gouvernement de Martin Hirsch. Le but de cette conférence est de présenter l'expérimentation sociale aux potentiels porteurs de projets ayant candidaté⁶⁴⁴. Le questionnaire central annoncé étant « *D'autres méthodes pour faire évoluer les politiques sociales ?* ».

Cette conférence organise une rencontre entre porteurs de projets, ou acteurs de terrain, personnes politique, partenaires sociaux et chercheurs en économie susceptibles de conduire des évaluations. Elle entérine le partenariat avec l'École d'Économie de Paris, le Crest et le J-Pal. Du fait de l'absence d'exemples français à mobiliser, l'argumentaire de promotion de cette méthode fait appel à la mobilisation d'alliés étrangers en invitant le Directeur du MDRC (Manpower Demonstration Research Corporation) – que nous avons évoqué dans notre premier chapitre - pour rendre compte des travaux américains et mexicains, discutés par François Bourguignon, directeur de l'École d'économie de Paris⁶⁴⁵. Les termes deviennent davantage techniques, on parle « d'expérimentations contrôlées » et de « tirage aléatoire », une table ronde sur l'état des savoirs sur les expérimentations sociales, en partenariat avec l'École d'Économie de Paris, donne la parole aux chercheurs promoteurs de cette méthode comme Eric Maurin (EHESS), Marc Gurgand (EEP), Bruno Crepond (Crest), William Pariente (JPAL Europe). Cette conférence propose aussi de revenir sur les premiers retours de l'expérimentation du contrat aidé, ou du RSA, avec l'Ansa. Elle organise enfin un Forum « meetic » des projets expérimentaux, où les porteurs de projets potentiels, qu'ils soient scientifiques ou acteurs de l'action publique ou associative, qui ont présenté une manifestation d'intention dans le cadre de l'appel à projets, présentent leurs travaux et échangent, en présence des membres du comité de sélection⁶⁴⁶.

⁶⁴² AGENCE NOUVELLE DES SOLIDARITES ACTIVES, « Revue succincte d'activité », 2006.2; LE HAUT COMMISSAIRE AUX SOLIDARITES ACTIVES CONTRE LA PAUVRETE, *Appel à projet d'expérimentation sociale 2008*, 2007.

⁶⁴³ LE HAUT COMMISSAIRE AUX SOLIDARITES ACTIVES CONTRE LA PAUVRETE, *op. cit.* p. 9.

⁶⁴⁴ *Lancement du Grenelle de l'insertion, les rencontres de l'expérimentation Sociale, programme*, 2007.

⁶⁴⁵ Voir le résumé de cette intervention sur le site <http://www.experimentationssociales.fr/vendredi.php>

⁶⁴⁶ Martin Hirsch revient sur cette démarche et explique qu'ils ont essayé de faire « le "Meetic" de l'expérimentation sociale, en disant : soit ce sont des chercheurs et des universitaires qui ont des protocoles à proposer et on essaiera de trouver des collectivités d'accord pour être terrain d'expérimentation ; soit on a des collectivités, des entreprises, des associations, des

L'objectif manifeste de cette conférence est d'expliquer aux porteurs de projets la nécessité de coupler leur projet avec la méthode d'évaluation expérimentale par assignation aléatoire. Un porteur de projet ayant participé à cette conférence explique qu'il était rendu explicite que le couplage du projet avec cette méthode d'évaluation, pratiquée notamment par des chercheurs renommés, augmenterait fortement leur chance d'être sélectionné et d'obtenir un financement. La création de cet appel à projets commence ainsi à constituer un champ d'action stratégique, dans lequel des ressources sont allouées, des positions sont en jeu, et la maîtrise de la méthode expérimentale par assignation aléatoire est favorisée par les règles du champ établies par les titulaires. On observe, par le lancement de cet appel à projets, un travail institutionnel important de la part du Haut Commissariat et des économistes précités, visant à favoriser le recours à l'expérimentation aléatoire. On note d'abord un travail de nature cognitive de définition de l'expérimentation sociale, couplant conduite de programme temporaire innovant et cette méthode d'évaluation. Ensuite, ce travail cognitif est assorti d'incitations financières, puisque le choix de cette méthode d'évaluation était perçu par les porteurs de projets comme un facteur déterminant dans l'accession au financement de leur projet.

Cet appel à projets peut être considéré comme un succès puisque 640 manifestations d'intentions ont été reçues⁶⁴⁷. Suite à une première sélection et des échanges avec les porteurs, 240 dossiers ont été examinés et finalement, le 30 janvier 2008, 37 projets ont été retenus. Cet appel à projets préfigure une reconfiguration de la nature des rapports entre science et action publique, puisque des équipes de chercheurs ont pu être à l'origine d'expérimentations sélectionnées. C'est le cas notamment du projet de l'équipe de Yannick L'horty, qui monte avec l'université Paris-Est Marne-la-Vallée (membre de sa fédération de recherche), le projet « Nouveau dispositif d'orientation en première année de licence ». Une université est porteuse, l'autre évaluatrice⁶⁴⁸. Le J-Pal propose 5 projets qui ont tous été acceptés. Se pose alors la question de l'acceptation du financement de projets de recherche dans le cadre de cet appel à projets, la réponse du jury est positive⁶⁴⁹. De fait, il est accepté que l'expérimentation devienne une nouvelle courroie de financement de laboratoires de recherche en France. Le champ scientifique, et les équipes universitaires connaissant une crise de financement, cette nouvelle source de fonds étatiques est une ressource inespérée. Quant au choix de la méthode d'évaluation à privilégier, Olivier Noblecourt commente le choix du jury dans son allocution lors d'un colloque organisé par la Dares : « *quels modes d'évaluation on veut privilégier ? Grandes querelles des sciences sociales, l'évaluation expérimentale est-elle la seule à devoir être reconnue par le jury ? Est-ce que nous admettons d'autres types d'évaluation ? Là-dessus, nous avons adopté une position commune, qui a consisté à privilégier très largement – et la constitution des dossiers de candidature y invitait fortement – la méthode expérimentale d'évaluation. Mais de ne pas s'y restreindre sur le principe.* »⁶⁵⁰

E. Maurin, membre du jury de sélection, explique qu'il a dû faire face à un écart entre ses critères et les projets proposés qui « *témoignaient d'un enthousiasme des acteurs,*

acteurs locaux qui ont des projets et, dans ce cas-là, on essaiera de trouver des chercheurs pour pouvoir les aider à les mettre dans un protocole qui rende les choses véritablement évaluables et dans des conditions dans lesquelles on peut en tirer des leçons DARES, Colloque sur les expérimentations pour les politiques publiques de l'emploi et de la formation, Paris, 22 mai 2008, p. 168, p. 8. Voir le résumé de cette deuxième journée sur le site <http://www.experimentationssociales.fr/samedi.php>

⁶⁴⁷ Olivier NOBLECOURT et Frédéric VALLETOUX, *L'expérimentation sociale locale en France: bilan, enjeux, perspectives*, Rapport de mission 2009-2010, rapport non publié, p. 55.

⁶⁴⁸ Un des évaluateurs de cette expérimentation explique lors d'un entretien que « *sur les sujets qui nous intéressent particulièrement on va monter des projets pour essayer de trouver des financements* »

⁶⁴⁹ Olivier Noblecourt « *Est-ce que notre rôle est de financer de la recherche universitaire stricto sensu ? On a finalement accepté de financer des projets qui étaient véritablement des projets de thèse, en réalité, ou pas loin* ». DARES, Colloque sur les expérimentations pour les politiques publiques de l'emploi et de la formation, op. cit. p. 135.

⁶⁵⁰ DARES, Colloque sur les expérimentations pour les politiques publiques de l'emploi et de la formation, op. cit. p. 136.

mais concevait l'évaluation comme demander aux bénéficiaires ce qu'ils avaient pensé ». Cet appel à projet « artisanal » est peu documenté, nous n'avons retrouvé ni trace de la composition des membres du jury, ni la liste des projets sélectionnés et de leur évaluation. Selon les communications des membres du jury, il semblerait que seulement environ 5 projets « *de professionnels qui s'étaient glissés dans le lot* » correspondaient aux critères scientifiques exigés⁶⁵¹. Ces projets-ci sont les plus documentés puisqu'ils font partie des 11 projets portant sur l'insertion des jeunes, dont la gestion fut reprise ensuite par le Fonds d'expérimentation pour la jeunesse⁶⁵². La mise en œuvre de ces projets est étudiée en détail dans le 3e chapitre de ce travail.

Au final, cet appel à projets a donc permis d'organiser le financement de cinq projets (sur 37), de petite taille, évalués par la méthode expérimentale par assignation aléatoire. Rapidement, de nouvelles initiatives émergent pour amplifier cet amorçage.

2.2. Le front scientifique : l'ancrage du champ d'action stratégique dans les champs de l'économie appliquée et de l'expertise d'État dans le secteur de l'emploi

En parallèle au travail du Cabinet politique pour organiser le lancement de projets ayant recours à cette méthode d'évaluation, d'autres initiatives sont conduites par les membres scientifiques de ce champ d'action stratégique. Elles visent à trouver des organisations souhaitant conduire et financer la méthode expérimentale notamment en diffusant un discours la légitimant. Cette recherche a déjà souligné plus haut, l'attrait que les économistes du Crest et de l'École d'Économie de Paris avaient pour la pratique de cette méthode. À partir de 2006, ces chercheurs multiplient les initiatives pour inciter les commanditaires d'évaluation à opter pour cette méthode, puisqu'elle ne peut être pratiquée sans leur aval. En tant qu'économistes appliqués, ils bénéficient déjà d'un accès privilégié aux champs de l'expertise de l'administration française qu'ils vont mobiliser pour convaincre les acteurs de l'action publique d'avoir recours à cette méthode. Par ailleurs, le travail institutionnel de constitution, d'organisation et de diffusion du discours autour de l'expérimentation social s'intensifie.

2.2.1. Les expérimentations dans le domaine de l'emploi, l'ancrage dans l'expertise d'État

Les chercheurs souhaitant pratiquer cette méthode ont besoin de démontrer qu'elle peut être appliquée en France, et sur des dispositifs de grande ampleur, leur donnant une puissance statistique suffisante. À la lumière des articles étrangers, ainsi que de leur première expérience, ils savent que l'appui du porteur du programme évalué, et son adhésion à la méthode est indispensable à la bonne conduite de l'expérimentation. Ils s'orientent ainsi vers les interlocuteurs de l'administration centrale qu'ils connaissent et dont la culture scientifique est suffisante pour trouver un écho attentif à leurs arguments.

Les chercheurs intéressés par la conduite d'expérimentations aléatoires sont des économistes appliqués qui ont déjà tissé des relations de confiance avec les commanditaires étatiques au cours de leurs collaborations précédentes. De plus, la plupart de ces chercheurs sont des administrateurs de l'Insee, ce qui leur confère un

⁶⁵¹ *Ibid.* p. 137-143.

⁶⁵² FEJ a récupéré la gestion par le biais d'un avenant aux conventions initiales actant le transfert de gestion de la DIIESES vers la DJEPVA FONDS D'EXPERIMENTATION POUR LA JEUNESSE, *Rapport d'activité 2009-2011*, 2012, p. 16.

statut hybride leur permettant d'alterner dans leur carrière, un travail de recherche au Crest avec celui d'administrateur dans les Services Statistiques Ministériels. De fait, on observe une porosité manifeste entre les membres du Crest et leurs potentiels commanditaires d'étude dans le domaine des politiques de l'emploi. Sur une très courte période, des chercheurs peuvent passer de prestataire à commanditaire comme le montrent les trajectoires de P. Zamora et de R. Rathelot qui ont conduit des expérimentations depuis une position interne à la Dares et ensuite externe quand ils sont passés au laboratoire du Crest.⁶⁵³ Un chercheur du Crest explique son trajet avant de rentrer dans ce laboratoire : *« à l'Insee on bouge tous les 3 ans, donc à la fin de mon poste à l'Insee, j'ai opté pour la Dares parce que l'un des contenus c'était de lancer, une des premières en 2007, grandes expérimentations sur les jeunes diplômés et c'est ça qui m'a décidé à choisir »*. D'ailleurs, les activités au sein de ces départements d'étude des ministères sont très proches des activités de chercheurs appliqués, étant donné qu'ils conduisent principalement des études en interne. Ainsi, dans les cas présentés ci-dessous, le choix d'avoir recours à cette méthode ne témoigne pas d'une volonté politique en amont, mais de la présence d'acteurs du champ d'action stratégique de l'expérimentation sociale en formation, partageant la même culture scientifique au sein même de l'administration centrale française, et particulièrement de ses Services de Statistiques Ministériels.

Les ingénieurs d'étude au sein de la Dares, ou des services internes de statistiques et études de l'ANPE (Agence Nationale pour l'Emploi), ont majoritairement une formation de statisticien, et ont suivi leurs études à l'Ensae (École nationale de statistique et d'administration économique), qui forme aussi bien des chercheurs que de futurs administrateurs. Ces milieux ne découvriraient donc pas cette méthode, puisqu'elle est présentée lors de leur formation universitaire, et que leur culture scientifique les a mis au contact d'articles étrangers fondamentaux dans le domaine de l'emploi mobilisant cette méthode⁶⁵⁴. Un scientifique ayant participé à la diffusion de cette méthode revient sur cette culture statisticienne au sein des corps d'État, propice à ce développement : *« on voit bien les corps qui ont envie de se saisir de ça (...), évidemment les administrateurs de l'Insee, donc là je crois qu'il y a un écho assez puissant dans l'administration, et puis plus ces corps avaient une formation scientifique plus l'écho était facile à avoir. »*

Cette méthode avait déjà été discutée notamment dans un document d'étude de 2005 de la Dares, mais concluant sur sa non-utilisation. Ce rapport la caractérisait d'évaluation *« complexe à mettre en œuvre et aux résultats limités »*, car peu reproductibles et soumises aux effets de déplacement lorsque l'on change d'échelle de mise en œuvre⁶⁵⁵.

C'est la conjonction de ce terreau de partage de normes, de pratiques, de constructions cognitives communes entre commanditaires et potentiels prestataires et, par ailleurs, l'initiative conduite par les entrepreneurs scientifiques et l'identification de fenêtres d'opportunités, qui a permis sa première utilisation de grande ampleur.

⁶⁵³ Un de ces chercheurs lorsqu'il revient sur son parcours a cette formulation explicite *« cette fois-ci je passe du côté recherche et je suis plus du côté ingénieur social »*

⁶⁵⁴ L'article de Lalonde plaçant pour l'utilisation de la méthode expérimentale en démontrant la différence entre les résultats obtenus avec une méthode quasi expérimentale a été souvent présenté lors de nos entretiens comme incontournable pour la plupart des statisticiens : Robert J. LALONDE, « Evaluating the Econometric Evaluations of Training Programs with Experimental Data », (1986) 76-4 *The American Economic Review* 604- 620..

⁶⁵⁵ Chirstine CHARPAIL, Tristan KLEIN et Serge ZILBERMAN, *Evaluation des politiques d'emploi: la deuxième génération des panels des bénéficiaires*, n°95, coll. Document d'études, Dares, 2005, p. 23.

2.2.1.1. L'expérimentation des opérateurs privés de placement

Depuis 2005-2006, les économistes du Crest et de l'école d'économie de Paris multiplient les initiatives afin de convaincre un porteur de projet de taille de commander le recours à cette méthode. À l'écart des initiatives du Haut Commissariat, ils savent saisir les opportunités se présentant à eux et effectuent un travail institutionnel intensif pour sa diffusion. Évoquant le travail d'un entrepreneur scientifique, un autre chercheur explique : *« il a vraiment une énergie telle qu'il a commencé à créer autour de lui un véritable élan »*, et *« il a tellement envie de faire des projets qu'il arrive à garder des relations institutionnelles avec beaucoup de gens »*. Ce travail a été notamment cognitif et consistait à présenter cette méthode comme la meilleure réponse aux problèmes politiques rencontrés. Ces opérations seront couronnées de succès puisque trois utilisations de cette méthode sont conduites dans le domaine de l'emploi, deux commanditées par Pôle Emploi et une autre par la DGEFP (Délégation générale à l'emploi et à la formation professionnelle).

En 2006, les services nationaux pour l'emploi français sont dans un état de crise et de recomposition de leur organisation et de leur modalités d'intervention. L'Unedic et l'Anpe, encore deux instances distinctes à l'époque avant de fusionner en Pôle Emploi, sont en porte à faux autour notamment de l'ouverture au recours à des opérateurs privés de placement pour l'accompagnement des demandeurs d'emploi. L'Unedic a expérimenté depuis 2005 un accompagnement renforcé par des opérateurs privés et clame qu'ils sont plus efficaces que les services de l'Anpe pour le retour à l'emploi des demandeurs d'emploi, sans pour autant disposer d'une évaluation de cette expérimentation. En réaction, l'Anpe décide de mener aussi une expérimentation d'accompagnement renforcé conduite par ses services (Cap vers l'entreprise), mais d'amener la preuve indéniable de son efficacité⁶⁵⁶. Dans cette configuration, comme nous l'avions envisagé dans notre première hypothèse, la méthode expérimentale est alors présentée par ses promoteurs scientifiques comme un instrument à même de résoudre ces tensions.

Afin d'évaluer ces expérimentations, l'Anpe se tourne vers ses interlocuteurs experts privilégiés ayant déjà conduit des évaluations pour cette institution dans ce domaine, - Marc Gurgand et Bruno Crepond - pour leur demander de réfléchir à un protocole d'évaluation. Ces derniers avaient déjà réalisé des études pour eux, notamment sur l'accompagnement des demandeurs d'emploi en 2005⁶⁵⁷. Ces chercheurs « persuadent » alors les commanditaires de l'évaluation, et notamment le directeur général des services, d'avoir recours à cette méthode afin de trancher le débat⁶⁵⁸. Cette démarche est relayée par leur réseau en interne, comme l'explique un des évaluateurs : *« Je connaissais un peu les gens du service statistique avec qui j'avais travaillé, je leur avais expliqué et c'est eux qui m'ont accompagné en interne, c'était bien, c'est eux qui ont trouvé aussi que c'était une bonne chose de faire ça, car eux aussi ils étaient confrontés à ce problème de l'absence d'évidence »*. Le choix de cette méthode est conforté par la

⁶⁵⁶ Une partie prenante de ce processus explique : *« Donc là, il y a eu un moment de tension très fort entre d'un côté l'assurance-chômage qui a dit « il faut ouvrir, mettre de l'émulation, etc. avec le service public lui-même, avoir recours à des opérateurs privés, on a nous-mêmes expérimenté un peu »*. D'ailleurs, ils l'avaient fait en 2005, dans des conditions qui n'ont jamais été tout à fait probantes, il n'y a pas eu de travaux très conclusifs là-dessus et ils avaient dit *« il faut absolument qu'on prouve que ça marche aux yeux de tous »*. Et donc, à ce moment-là, l'ANPE a dit *« chiche, mais nous on va mettre en place un dispositif internalisé d'accompagnement renforcé, avec nos propres compétences, et on verra qui fait le mieux »*. ».

⁶⁵⁷ Bruno CRÉPON, Muriel DEJEMPEPE et Marc GURGAND, « Counseling the Unemployed: Does It Lower Unemployment Duration and Recurrence? », *discussion paper series IZA* 2005.1796.

⁶⁵⁸ Ce terme est employé par Claude Siebel président du comité de pilotage de cette évaluation lors d'une journée organisée par la SFE, Claude SIEBEL, François AVENTURE et Stéphane LARDY, *L'évaluation de l'accompagnement renforcé des demandeurs d'emploi*, Journée de la S.F.E., Cheminer dans l'évaluation, 6 avril 2010. Il est corroboré par l'un des évaluateurs qui déclare lors d'un entretien *« quand ils nous ont contacté (Pôle Emploi) on leur a dit : c'est ça qui faut faire »*

décision de l'Unedic de soumettre aussi son dispositif d'accompagnement renforcé conduit par les opérateurs privés aux mêmes critères d'évaluation, afin de trancher une fois pour toutes le débat entre l'efficacité du public ou du privé dans l'accompagnement des demandeurs d'emploi.

Il fallait donc trouver la démarche la plus sûre pour estimer l'impact ainsi que les effets différentiels des deux dispositifs concurrents. Le recours à cette méthode a donc été choisi suite à la proposition des chercheurs du Crest, avec le soutien de la Dares et, elle a été présentée comme la seule à même d'apaiser le débat, par l'administration d'une preuve indéniable de l'impact sur le retour à l'emploi de ces deux dispositifs. D'ailleurs le recours aux services de ces chercheurs ne s'effectue pas dans le cadre de marchés publics, mais de subventions de projets scientifiques avec le Crest et l'École d'Économie de Paris. Ce mode de financement par subvention permet de contrôler l'entrée dans le champ d'action stratégique en restreignant l'accès aux ressources aux titulaires du champ qui par leur liens privilégiés avec l'administration centrale peuvent intervenir en amont sur le choix méthodologique. Ce mode de subvention était justifié par une démarche scientifique d'accumulation de savoir scientifique sur des thèmes d'intervention de politique publique⁶⁵⁹. On retrouve ici la même dynamique présente dans l'appel à projet de la Diieses qui, passant par la méthode expérimentale, ouvre de nouveaux modes de financement par subvention d'une recherche appliquée par l'État, mais néanmoins valorisable par les scientifiques dans le champ scientifique. Ce positionnement des économistes statisticien appliqué à l'intersection de la sphère de l'action publique et de la sphère scientifique leur permet donc d'une part, d'agir sur le champ de l'action publique pour l'influencer dans ces choix méthodologiques et d'autre part, de garder un statut d'expert externe de part leur rattachement à des laboratoires de recherche.

Est alors créé un comité d'évaluation *ad hoc*, représentant les différentes parties pour répondre à un besoin de pacification des relations entre Unedic et Anpe, présidé par un ancien directeur de la Dares très favorable, de prime abord, à l'utilisation de cette méthode, comme en témoigne une de ses déclarations citée par un de nos interlocuteurs : *« moi, toute ma carrière j'aurais voulu faire de l'évaluation par comparaison de groupe, et c'était interdit et donc, dès qu'on me l'a proposé, bien sûr, j'ai été d'accord pour y participer et présider l'instance »*. Selon les parties prenantes rencontrées, la Dares a joué le rôle de courroie de transmission, de traducteur entre les exigences politiques et les impératifs scientifiques.

Ce dispositif représente la première utilisation de cette méthode à cette échelle et avec rigueur, bien que son emploi ait pu être source de conflits (ce cas est étudié plus en détail dans le chapitre 3). Les chercheurs ont donc leur premier exemple français mobilisable pour leur argumentaire de démonstration de la faisabilité de cette méthode sur le territoire français. Néanmoins, cet exemple ne s'inscrit pas dans une démarche d'*evidence-based policy* puisqu'il apparaît rapidement que le résultat n'a pas pour but d'influencer la décision publique d'avoir recours à des opérateurs privés, mais tient de la démonstration scientifique et de la démonstration du savoir faire de l'action des services publics pour l'emploi.

⁶⁵⁹ Un interlocuteur de Pôle Emploi précise « En l'occurrence, aussi bien sur l'opération qui a porté sur l'accompagnement renforcé des demandeurs d'emploi que sur celle plus récente qui a eu trait au CV anonyme, on a travaillé avec le CREST et l'école d'économie de Paris dans un cadre qui était un cadre non pas de marchés publics, mais de subventions en projet scientifique. Puisqu'ils étaient eux dans une perspective d'accumulation et de recherche sur à la fois les questions d'efficacité des politiques d'emploi et, pour le deuxième thème, de discrimination sur le marché du travail, et donc on a pu justifier de la sorte qu'on puisse subventionner l'équipe de recherches pour mener avec nous une opération d'intérêt commun. »

Dans le cadre de cette évaluation, dont l'initiative vient des services, et non d'une commande politique, il n'a nullement été précisé que le résultat de cette expérimentation conditionnerait le recours à des opérateurs privés de placement ou non. Le recours aux opérateurs privés était perçu comme une opération nécessaire pour soulager les équipes de conseillers qui ne pouvaient faire face aux besoins d'accompagnement des demandeurs d'emploi. Un interlocuteur de Pôle emploi résume, « *on n'est pas du tout dans une logique d'expérimentation sociale, même si on emprunte aux méthodes qui s'y rapportent* ». Cette initiative profite surtout de l'opportunité du déploiement progressif d'un nouveau dispositif, pour tester son efficacité et apporter des « *arguments pour éclairer le débat politique* », mais non dans l'idée de généraliser le dispositif le plus efficace à la lumière des résultats.

En effet, comme le souligne Bruno Crépon dans la presse, il s'avère que *"le principal enseignement est que la plus-value apportée par l'ANPE avec son programme est nettement supérieure à celle des opérateurs de placement privé dans les mêmes bassins d'emploi, et sur la période d'observation la plus pertinente, c'est-à-dire après six mois d'accompagnement"*, alors que Pôle Emploi entérine le recours à des opérateurs privés de placement⁶⁶⁰. Mais au-delà de la présentation des résultats, les évaluateurs insistent sur l'importance d'avoir utilisé pour la première fois la méthode d'évaluation expérimentale, comme le montre le même article de presse dans lequel le chercheur ajoute : *"C'est la première fois que des méthodes économétriques d'évaluation des politiques publiques sont réellement utilisées en France"*. Cet exemple atteste ainsi d'un premier développement de cette méthode scientifique, tout en restant à l'écart d'une démarche d'*evidence-based policy*, du fait, en partie, de l'absence de commande politique à son origine, mais aussi de ses conclusions allant à l'inverse de la tendance d'ouverture de l'accompagnement des demandeurs d'emploi par des opérateurs privés. Alors que cette méthode a été présentée comme une ressource scientifique dont la rigueur permettrait de trancher les débats politiques, il semble qu'au final, elle n'a pour l'instant pas d'échos dans la programmation politique. Cette initiative témoigne néanmoins du succès de ces entrepreneurs à permettre l'émergence du champ d'action stratégique de l'expérimentation sociale en cherchant à mobiliser le champ de l'expertise d'État dans le domaine de l'emploi, lui-même en crise de légitimité et en quête de nouveaux outils d'expertises.

2.2.1.2. L'expérimentation scientifique au cœur de l'administration centrale, l'expérimentation « Jeunes diplômés »

À la même période, en début d'année 2007, avant la nouvelle présidence de la République, les chercheurs continuent la diffusion de cette méthode par le vecteur de la Dares avec une nouvelle expérimentation sur le recours aux opérateurs privés de placement, mais cette fois pour les jeunes diplômés. Cette initiative émerge cette fois-ci directement du cabinet du ministre, elle est financée par la DGEFP (délégation générale à l'emploi et à la formation professionnelle), laquelle désigne la Dares pour conduire l'évaluation. Le cabinet veut évaluer cette expérimentation, mais n'a pas en tête une méthode d'évaluation en particulier⁶⁶¹. Deux administrateurs de la Dares souhaitent saisir cette opportunité pour utiliser la méthode aléatoire, ils sollicitent alors le soutien de scientifiques renommés (Bruno Crépon et Esther Duflo) pour appuyer cette démarche. Cette démarche de promotion est assumée par un des évaluateurs lors d'un

⁶⁶⁰ Jamila ARIDJ, « L'accompagnement renforcé plus efficace pour les chômeurs », *Le point* (4 juillet 2008).

⁶⁶¹ Source, entretien avec une des parties prenantes de la Dares

entretien : « *Le travail que fait la Dares c'est aussi de sélectionner et de favoriser l'adoption raisonnée d'une nouvelle méthode.* ». Ces scientifiques répondent présents, et veulent aussi participer au cadrage de cette expérimentation afin de répondre à des questionnements scientifiques. Un de nos interlocuteurs de la Dares explique : son « *aide a été centrale puisqu'elle (Esther Duflo) a dit oui « il faut absolument faire ça et en plus il faut évaluer les externalités* » ».

En effet, cette étude est d'un intérêt particulier pour les chercheurs, d'abord du point de vue des moyens, puisqu'elle permet une enquête en quatre vagues sur 35 000 jeunes avec un budget d'environ 600 000 euros⁶⁶². Mais surtout, elle revêt un intérêt scientifique particulier puisque son ingénierie permet de renforcer la validité externe de ses résultats en prenant en compte les effets de déplacement, en procédant à un double tirage aléatoire des zones d'expérimentations et ensuite des personnes accompagnées⁶⁶³. Cette modalité de tirage au sort permet de savoir si les politiques d'accompagnement renforcées par des opérateurs privés ont un effet, mais surtout si elles ne réduisent pas les chances des gens qui n'en bénéficient pas de trouver un emploi, en changeant seulement les places dans la file d'attente pour l'emploi. L'équipe de scientifique est parvenue ici à prendre la main sur l'ingénierie de l'expérimentation en décidant en partie de l'objet de l'évaluation. De plus, la centralisation de cette expérimentation au sein de la Dares a permis un fort contrôle du protocole.

Néanmoins, ces efforts de rigueur scientifique ne rencontrent pas l'intérêt politique. Bien qu'elle soit issue d'une commande politique, peu à peu cette étude n'est plus suivie par le cabinet. Cette perte d'intérêt s'explique en partie par le temps nécessaire à la mise en place de l'étude qui a pris six mois entre sa commande et son lancement en septembre 2007, et au temps de production et d'interprétation des résultats, le premier papier ne paraissant que quatre ans plus tard, laps de temps suffisant pour voir passer trois ministres de l'emploi différents⁶⁶⁴. Une autre explication vient de la surprise des chercheurs devant les premiers résultats produits, qu'ils peinent à interpréter, puisqu'ils concluent à une non-effectivité du dispositif et à l'absence d'effets externes. Un des évaluateurs explique : « *on avait un problème parce qu'on disait qu'il n'y avait pas d'effet, pas d'effet externe et on était compl... enfin... bon c'était un peu compliqué parce qu'on se disait, c'est bizarre, c'est vraiment bizarre parce que la théorie quand même elle est claire* ».

De fait, de prime abord, les résultats de cette expérimentation ne s'avèrent pas être un allié mobilisable, ni par le politique puisqu'ils concluent à l'ineffectivité du dispositif, ni par les scientifiques, puisque leur interprétation s'avère difficile. Ce n'est que plus tard, en retravaillant les résultats sur des sous ensembles de population qu'ils observeront des effets, mais qui restent politiquement difficiles à formuler puisque, selon l'un des évaluateurs rencontrés, les leçons politiques à en tirer sont que l'accompagnement renforcé à plus d'effet lorsqu'il y a beaucoup de chômage, mais au détriment des autres personnes sur la file d'attente, alors qu'en période de faible taux de chômage, il a un effet moindre, mais ne défavorise pas les autres demandeurs d'emploi. Notre interlocuteur commente : « *C'est quand même très intéressant pour la conduite de*

⁶⁶² Le montant de l'évaluation a été précisé par une des parties prenantes, mais aucune trace écrite n'a été retrouvée, le protocole est présenté dans Bruno CREPON, Esther DUFLO, Marc GURGAND, Roland RATHELOT et Philippe ZAMORA, « L'accompagnement des jeunes diplômés demandeurs d'emploi par des opérateurs privés de placement: les enseignements d'une évaluation », *Dares Analyses* 2011.94.

⁶⁶³ Les effets de déplacement ou de « file d'attente » sont très débattus dans la littérature des dispositifs de retour à l'emploi, et remettent en cause la capacité de certaines expérimentations à prendre en compte l'effet de la généralisation d'un dispositif.

⁶⁶⁴ Bruno CREPON, Esther DUFLO, Marc GURGAND, Roland RATHELOT et Philippe ZAMORA, « L'accompagnement des jeunes diplômés demandeurs d'emploi par des opérateurs privés de placement: les enseignements d'une évaluation », *Dares Analyses* 2011.94.

la politique économique, parce que bien évidemment c'est quand le marché du travail va mal qu'on dit qu'on va aider les gens. Le problème c'est que ce résultat ne peut être publicisé, puisqu'il n'est pas populaire politiquement, puisqu'il montre que ces choix (d'aider davantage les demandeurs d'emploi en temps de crise) ne tiennent pas de l'efficacité économique, mais de la redistribution économique. »

Ainsi, dans ce cas se dessine un écart entre la perspective cumulative de savoir sur le long terme des scientifiques et l'immédiateté du champ politique. Le financement de cette étude s'inscrit aussi à la frontière du financement de recherche scientifique puisque la Dares savait à quels chercheurs elle souhaitait confier l'évaluation, et que cette dernière a pu être orientée afin de servir les questions théoriques des évaluateurs/chercheurs.

2.2.1.3. L'expérimentation du CV anonyme, première incarnation d'une rhétorique *d'evidence-based policy*

On retrouve un autre exemple de collaboration de ces chercheurs avec Pôle Emploi en 2009, par l'évaluation du CV anonyme, financée aussi par une subvention à la recherche. Selon nos interlocuteurs, ces subventions permettent de « *faciliter la vie du chercheur* », de les aider à financer des colloques, de payer des doctorants pour travailler sur ces missions. Dans le cas du CV anonyme, la démarche d'*evidence based policy* est manifeste. La loi du 31 mars sur l'égalité des chances précisait, dans un de ses articles, que toute entreprise de plus de 50 salariés devait procéder au recrutement de ses salariés par CV anonyme. Néanmoins, le décret d'application n'a pas été promulgué. Le commissaire à la diversité et à l'égalité des chances, Yazid Sabeg, commande alors une étude pour tester l'effectivité de cette mesure et décide de prendre la décision de promulguer le décret ou non sur la base des résultats. Cet exemple incarne typiquement la subordination d'une loi votée démocratiquement par le parlement à un résultat d'évaluation d'impact produit par des scientifiques. Cette évaluation est confiée à Pôle Emploi qui décide de mobiliser la même équipe de chercheurs du Crest et d'avoir recours à la méthode randomisée puisque, comme le souligne un interlocuteur de cette organisation, c'est « *en théorie* », la méthode la plus puissante. Pôle Emploi assume cet engagement pour la promotion de cette méthode et du travail de ce groupe de chercheurs : « *ça s'inscrit aussi dans le cadre général de J-PAL, avec Esther Duflo, et donc leur stratégie est assez expansionniste, ça consiste à dire « nous avons là un collectif de chercheurs qui intervient dans différents domaines et dans différents pays, et qui est dans une logique qui est, en Europe en tout cas, de développer la présence de J-PAL, notamment par le biais d'une entrée vers les services publics de l'emploi ». Donc c'est tout à fait clair, ils ne s'en cachent pas, et nous on a essayé de le favoriser* ».

Cependant, le protocole ne peut être mené dans les règles de l'art, et les chercheurs ne peuvent tirer au sort les entreprises entrant dans l'expérimentation, biaisant ainsi la représentativité de l'échantillon. Finalement, l'étude conclut à une non-effectivité du CV anonyme, et même parfois, à un effet négatif de ce dispositif sur l'employabilité des jeunes défavorisés⁶⁶⁵. Néanmoins, les scientifiques modèrent ces résultats, en estimant qu'ils peuvent en partie tenir aux biais d'un échantillon composé d'entreprises volontaires, qui seraient plus favorables à la lutte contre les discriminations par rapport à l'ensemble des entreprises sur le territoire, mais aussi du manque d'information pour que l'employeur puise interpréter les informations du CV. Mais ce résultat, allant dans le sens de la position du gouvernement, est utilisé pour légitimer la décision de ne pas

⁶⁶⁵ Luc BEHAGHEL, Bruno CREPON et Thomas LE BARBANCHON, *Evaluation de l'impact du CV anonyme, synthèse et rapport final*, PSE, CREST, J-Pal, 2011.

promulguer le décret. Cette méthode, non exercée dans les règles de l'art, peut difficilement aboutir à des résultats tranchés sur l'efficacité d'un dispositif, et mène à la production d'un discours nuancé, peu audible dans le champ politique, et sujet à une instrumentalisation échappant aux producteurs de ces travaux.

Ces premières incarnations d'expérimentations par assignation aléatoire dessinent une reconfiguration de l'interaction entre action publique et sciences sociales. En effet, avec l'évaluation aléatoire émerge une nouvelle source de financement de la recherche, à mi-chemin entre évaluation et travaux d'économie appliquée. Cette confusion avait déjà été anticipée par des observateurs des années 1990, qui ont estimé que la réduction des crédits de l'État pour la recherche tendrait probablement à une confusion progressive des lignes budgétaires consacrées à l'évaluation de l'action publique et à la recherche⁶⁶⁶.

2.2.2. L'intensification du travail institutionnel pour stabiliser le champ d'action stratégique

La période de 2007 à 2009 est aussi un moment d'intensification du travail cognitif et de diffusion de la promotion de cette méthode scientifique comme étant la plus rigoureuse et permettant d'instaurer les conditions de *l'evidence-based policy*. Cette promotion est organisée notamment par la création d'une nouvelle institution scientifique, le J-Pal (Abdul Latif Jameel Poverty Action Lab) Europe dédiée à cette activité. Par ailleurs les acteurs des différentes initiatives de promotion se retrouvent pour s'unir autour d'un discours commun.

2.2.2.1. La création du J-Pal Europe à Paris

Le J-Pal est un réseau de chercheurs créé en 2003 aux États-Unis et dédié à la pratique de la méthode d'évaluation aléatoire (voir chapitre 1). Il présente l'évaluation aléatoire comme une nouvelle démarche offrant la meilleure méthode scientifique pour évaluer les programmes sociaux et démontrer objectivement qu'un projet remplit les objectifs promis⁶⁶⁷. La décision de l'implantation française répond à l'atteinte d'une masse critique de conduites de projets expérimentaux nécessitant un appui structuré. Les origines françaises d'Esther Duflo ont eu aussi un poids non négligeable dans cette décision⁶⁶⁸. Cette dernière délaisse d'ailleurs pour un semestre son poste au MIT afin d'appuyer l'Ecole d'Economie de Paris, au printemps 2007, dans la création du J-Pal Europe. Sa création, en mai 2007, a pour objectif de « *soutenir les projets de l'organisation basés en Europe et accroître la participation des chercheurs européens* »⁶⁶⁹.

Le J-Pal Europe est à l'époque un réseau d'une dizaine de chercheurs dont cinq Français. Il a pour but de faciliter la conduite des évaluations expérimentales, en proposant des formations à sa conduite et en organisant aussi la diffusion des résultats par la préparation de matériel de communication ou le dialogue avec des ministères⁶⁷⁰. La première formation, proposée en mai 2008, est un succès puisqu'elle a réuni plus de

⁶⁶⁶ Jean-Noël CHOPART et Claude MARTIN, *Sciences de l'Homme et de la société et évaluation des politiques publiques, Evaluation et Décision, séminaire de méthode*, coll. Plan Urbain, CEOPS, Ecole Nationale des Travaux Publics de l'Etat, 19 septembre 1990, p. 139- 149, p. 144.

⁶⁶⁷ J-PAL, *Une nouvelle démarche : l'évaluation aléatoire*.

⁶⁶⁸ Un des scientifiques proches de la chercheuse explique « elle voulait créer en France depuis longtemps un laboratoire qui permet de couvrir l'Europe et aussi qui soit un point de départ facile pour de départ pour les pays d'Afrique francophone », La fondation du J-Pal Europe est expliquée par Esther Duflo dans DARES, *Colloque sur les expérimentations pour les politiques publiques de l'emploi et de la formation*, op. cit. p. 161.

⁶⁶⁹ Source : <http://www.povertyactionlab.org/fr/europe>

⁶⁷⁰ Les cinq Français sont : Luc Behagel, Bruno Crépon, Marc Gurgand, Eric Maurin et Philippe Zamora, et par ailleurs William Parienté a participé à des expérimentations françaises lorsqu'il était post-doc pour le J-Pal

43 personnes. Le J-Pal Europe offre aussi les moyens de conduire des expérimentations en organisant notamment le travail d'enquête par le recrutement et la coordination d'assistants de recherche. Cette logistique est estimée comme nécessaire par certains scientifiques rencontrés, qui considéraient ne pas avoir l'expérience suffisante pour effectuer des études de cette ampleur. Il permet aussi une mise en réseau, comme le souligne un membre du J-Pal : *« on partage un peu le même point de vue, c'est une opportunité de travailler avec ces gens, vous voyez M. je la connaissais, mais pas très bien donc là ça m'a amené à travailler avec elle sur cette initiative. Ca apporte aussi parfois... je ne sais pas, je pense qu'on est plus fort quand on doit convaincre quand on a une démarche collective, ce n'est pas garant de crédibilité, mais ça n'enlève rien. »*. Le J-Pal Europe a aussi vocation à servir d'intermédiaire entre les scientifiques souhaitant conduire cette méthode et les porteurs de projets désireux de mettre en œuvre des expérimentations.

Le J-Pal Europe est donc une institution dédiée intégralement à la promotion de l'utilisation de la méthode expérimentale, en offrant à ses membres un appui logistique à la conduite de cette méthode pour faciliter sa pratique et lever les barrières qui pourraient les décourager de s'inscrire dans cette démarche. L'appartenance à ce groupe offre un accès privilégié au statut de *titulaire* dans le *champ d'action stratégique* de l'expérimentation sociale. À partir de 2007, les chercheurs ne sont donc plus dans une conduite « artisanale » de la méthode, et ils bénéficient de cet appui organisationnel pour gérer des équipes, ainsi que d'un appui relationnel entre scientifiques, et aussi communicationnel en direction de la société civile et des acteurs publics. Cette structure est accueillie par l'École d'Économie de Paris et lui est rattachée officiellement en mars 2008, ce qui contribue à conforter l'investissement de ses chercheurs pour la promotion de cette méthode. L'accueil du J-pal Europe s'inscrit dans la stratégie de l'École d'Économie de Paris de diversifier les sources de financements en ayant recours, notamment, aux financements par les marchés publics ou de subventions d'Etat, et dans sa stratégie d'augmentation de la visibilité de l'école par l'inscription dans un réseau international. La pratique d'expérimentations sociales répond donc à un besoin de financement, ainsi qu'à une volonté de favoriser la visibilité des économistes français sur la scène internationale.

La mise en place de cette structure, ainsi que les différentes initiatives, témoignent de la présence et de la structuration institutionnelle d'une offre scientifique voulant pratiquer cette méthode. Néanmoins, la demande du côté des commanditaires reste encore faible, et surtout, les premières expérimentations menées ont montré la nécessité d'une acclimatation culturelle à cette démarche, notamment par le dépassement des barrières éthiques ancrées culturellement au vu du gain de rigueur scientifique spécifique à cette méthode. L'année 2008, témoigne d'une stratégie d'intéressement par la diffusion de cadres cognitifs auprès des acteurs scientifiques, politiques et administratifs, par le biais notamment de la mobilisation d'alliés sous la forme de présentation d'expérimentations conduites en France ou à l'étranger, et de scientifiques renommés apportant cette légitimité à cette démarche.

2.2.2.2. Le colloque sur les expérimentations pour les politiques publiques de l'emploi et de la formation organisé par la Dares

La Dares affirme son soutien au développement de cette méthode par l'organisation d'un colloque consacré à la promotion de cette méthode, en partenariat avec le J-Pal. Ainsi, le 22 et 23 mai 2008, la Dares, « avec l'appui décisif » d'Esther Duflo et de Bruno Crépon, organise à Paris un « *colloque sur les expérimentations pour les politiques publiques de*

l'emploi et de la formation »⁶⁷¹. Ce colloque illustre particulièrement bien le discours de légitimation de la méthode expérimentale, ainsi que la composition du *champ d'action stratégique* la défendant. Derrière le terme expérimentation, il faut ici comprendre la méthode scientifique aléatoire, puisque ce colloque y est entièrement dédié. Seule une intervention présente l'expérimentation normative⁶⁷². Toutes les interventions la mentionnent, voire y sont consacrées, et les plus grands spécialistes mondiaux sont présents. L'existence d'actes de ce colloque offre un accès privilégié au travail de problématisation opéré à cette occasion. Au vu de la teneur scientifique des débats, il est clair que ce colloque s'adresse à la communauté scientifique. Les membres de la Dares, comme par exemple Beatrice Sédillot, la commanditaire de l'expérimentation « *jeunes diplômés* » évoquée plus haut, président les séances.

Martin Hirsch explicite, en introduction, la stratégie de promotion de cette méthode par son cabinet, et revient sur les éléments centraux de la problématisation. Face à un auditoire scientifique, il insiste ici sur les aspects scientifiques de son identité en critiquant les modes d'action de l'État français ou des acteurs politiques. On retrouve son inscription dans le mouvement de *evidence-based policy*, il met en cause « *une sorte de tradition française d'avoir la science infuse plutôt que d'y aller par des méthodes de recherche, de politique fondée sur l'évidence (...) au moment où se font les discussions des grands choix politiques, éviter qu'elles se fassent dans le vide ou de manière idéologique, mais qu'elles puissent se fonder sur des données expérimentales solides* »⁶⁷³. Mouvement nécessaire, au vu de l'importance des dépenses publiques dans le domaine social et de la demande de retour sur investissement. La « *force de l'évidence* » permet, selon lui, de légitimer les dépenses dans le social. Il justifie son souhait de conduire des programmes expérimentaux, et son travail avec les « universitaires français » comme Marc Gurgand et Eric Maurin pour investir dans les travaux scientifiques français expérimentaux. Il développe cette logique de promotion et de constitution du champ d'action stratégique : « *On s'est habitué à contourner, ruser avec les différents obstacles, pour que la logique de l'expérimentation soit introduite dans ce pays et sur une politique sociale importante* »⁶⁷⁴. Il affirme son projet de manœuvrer juridiquement pour que des dépenses publiques soient réservées au financement d'expérimentations, ainsi que la mobilisation de fonds privés et de fondations.

Les plus grands spécialistes de cette méthode scientifique sont là, dont Esther Duflo, Abhijit Banerjee et Joshua Angrist du MIT. Ils consacrent leur présentation à la défense de la méthode aléatoire en mobilisant des exemples d'applications étrangères comme *Progressa* au Mexique ou le *Self Sufficiency Project* au Canada. La teneur des débats de la première journée est hautement scientifique dans la démonstration de la supériorité de cette méthode pour estimer l'impact d'une intervention. Celle-ci est d'ailleurs présentée comme permettant de pallier à l'absence de loi du marché, du fait de l'absence de compétition en politique publique⁶⁷⁵. Sa force politique est réaffirmée, puisque cette méthode « *incroyablement simple* », sans « *statistiques élaborées* », permet de produire

⁶⁷¹ DARES, *Colloque sur les expérimentations pour les politiques publiques de l'emploi et de la formation*, op. cit. p. 4.

⁶⁷² Voir « Histoire et perspectives du cadre juridique des expérimentations législatives et réglementaires en France » Gwénaële Calves, Université de Cergy-Pontoise

⁶⁷³ DARES, *Colloque sur les expérimentations pour les politiques publiques de l'emploi et de la formation*, op. cit. p. 5-9.

⁶⁷⁴ DARES, *Colloque sur les expérimentations pour les politiques publiques de l'emploi et de la formation*, op. cit. p. 8.

⁶⁷⁵ Abhijit Banerjee "What is distinct about a lot of social policy is that for two distinct reasons there is no market test. One I think because there is often no choice, you have to send your children to school, and there is only one school, if the school doesn't work then you still just send your child to that school. That's one reason. The other reason is that governments often want to make you do things that you don't want to do."p12

des résultats évidents ne pouvant être contredits⁶⁷⁶. L'un de ces scientifiques résume l'enthousiasme des intervenants de cette journée par : « *Nous sommes comme des enfants ayant trouvé un nouveau jouet*⁶⁷⁷ ».

Des exemples français sont aussi présentés dans la deuxième journée, notamment par Bruno Crépon. Il souhaite illustrer le tirage aléatoire en pratique, afin de démontrer que « *l'échantillonnage aléatoire n'est pas une idée bizarre, mais que c'est plutôt un principe général qui peut se décliner et s'insérer très facilement, ou en tout cas de façon possible, dans les contraintes opérationnelles de la conduite d'un projet.* »⁶⁷⁸. Il mobilise pour cela les exemples de l'expérimentation dans les Hauts-de-Seine et l'expérimentation de Pôle Emploi et de la DGEFP sur le recours à l'accompagnement renforcé et aux opérateurs privés. Il explicite aussi l'entreprise scientifique commune : « *nous sommes un ensemble de chercheurs, principalement du CREST et de l'École d'économie de Paris, qui essayons de développer ces méthodes d'évaluation depuis plusieurs années.* »⁶⁷⁹.

2.2.2.3. Les rencontres de l'expérimentation sociale en Europe

Dans la même année, le Haut Commissariat aux Solidarités Actives contre la pauvreté étend ses perspectives d'intéressement avec l'organisation des « rencontres de l'expérimentation sociale en Europe », les 21 et 22 novembre 2008, à Grenoble, dans le cadre de la présidence française de l'Union Européenne. Alors que le colloque de la Dares s'adressait à un auditoire scientifique, celui-ci mobilise les acteurs de l'action publique, soit les potentiels commanditaires ou porteurs de projets. On constate à cette période une dissociation de la mobilisation du champ scientifique d'une part, et du champ de l'action publique d'autre part, permettant d'ajuster le discours de promotion de l'expérimentation vers son pendant scientifique auprès du premier, et vers son pendant politique au sens d'innovation sociale auprès du second. On observe ici clairement, une démarche d'entrepreneur frontière ajustant son discours selon les intérêts et cadres cognitifs des supports qu'il souhaite rallier. Cette conférence fait écho à un travail, amorcé en mars 2008, de sollicitation de l'Union Européenne au sujet de l'expérimentation sociale. Le ministère français des Affaires étrangères et européennes en vue de la prochaine présidence française de l'Union Européenne, à la demande du haut commissaire aux solidarités actives avait demandé au Comité Economique et Social Européen d'élaborer un avis exploratoire sur la question suivante « Comment l'expérimentation sociale peut-elle servir en Europe à l'élaboration des politiques publiques d'inclusion active ? ». L'avis adopté à l'unanimité (66 voix pour et 1 abstention) est ambiguë puisqu'il soutient le développement de l'innovation sociale tout en avertissant sur son couplage avec la méthode expérimentale aléatoire ainsi que son instrumentalisation potentielle par les acteurs politiques.

⁶⁷⁶ Abhijit Banerjee « *is incredibly simple* » » the advantage of that is that it is hard to argue with the results, you know where it came from, you know what exactly was done. So as a result, politically it is often extremely powerful because you cannot say, well you did this statistical technique to control for this or that, or the way you did it was not the right way to analyze the data, and if I analyze the data in some other way, we'll get some other conclusion"(...) "no fancy statistics"(...) "The results are almost as they are, you see them or you don't see them." p12 Ibid.

⁶⁷⁷ Traduction de l'auteur de « *We are sort of like children who have found a new toy.* » Ibid p19

⁶⁷⁸ DARES, Colloque sur les expérimentations pour les politiques publiques de l'emploi et de la formation, op. cit. p. 124.

⁶⁷⁹ DARES, Colloque sur les expérimentations pour les politiques publiques de l'emploi et de la formation, op. cit. p. 125.

Avis du Comité économique et social européen sur « Comment l'expérimentation sociale peut elle servir en Europe à l'élaboration des politiques publiques »

Bien que le comité soit convaincu de l'utilité primordiale de l'expérimentation innovante comme outil d'élaboration des politiques publiques, il émet quelques réserves quant à son utilisation.

Ce comité détache la conduite de politiques innovantes et leur évaluation par la méthode aléatoire, celle-ci n'est mentionnée qu'une fois et c'est pour l'exclure de la définition de l'expérimentation sociale.

- Ils estiment que l'expérimentation sociale est une nébuleuse conceptuelle et que les évaluations sont souvent trop vagues (2.1.2).
- Ils sont « choqués » par le terme expérimentation (3.2.1.1.) car « les êtres humains ne sont pas des cobayes » et proposent de parler simplement d'expériences ou d'innovations.
- Ils tentent alors à dissiper autant que possible les zones de « flou » dudit concept et préfèrent la définition proposée par le Haut Commissariat (une innovation de politique sociale, initiée dans un premier temps à petite échelle, compte tenu des incertitudes relatives à ses effets, mise en œuvre dans des conditions qui permettent d'évaluer ceux-ci, et dans l'optique d'une généralisation ultérieure) que celle de l'Urban institute (l'affectation aléatoire des bénéficiaires et d'un groupe de contrôle, un dispositif de politique publique, un dispositif de suivi une évaluation). Ils excluent ainsi le couplage systématique de l'expérimentation avec son évaluation aléatoire.
- Ils émettent des réserves sur l'utilisation des expérimentations pour éclairer l'action publique :
« 3.2.1.3 Leurs spécificités de lieux et d'échelle rendent les expérimentations généralement et inévitablement non reproductibles »;
« 3.2.1.4 Les expérimentations risquent de servir d'alibis pour des décideurs publics peu enclins à réaliser des réformes générales. Elles peuvent conduire à réduire, voire abolir, des dispositifs sociaux existants et protecteurs; »

Ils annoncent des actions de promotion de la démarche d'innovation sociale mais sans en préciser la méthode d'évaluation :

« 4.1.1 le 2 juillet 2008, la Commission a adopté l'Agenda social renouvelé, qui comprend une importante communication sur le renforcement de la Méthode ouverte de coordination sociale. Le texte souligne que Progress soutiendra « les expérimentations sociales ».

« 4.3.1 il préconise avec insistance la constitution d'un dispositif en forme de réseau européen attentif, ayant mission de développer et partager la connaissance de l'existence, la nature, la teneur, les modalités, les enseignements, les résultats des expérimentations accomplies dans les états de l'Union. »

« 4.3.4 Le Parlement et le Conseil pourraient instaurer en commun de façon systématique au moins une rencontre annuelle de « l'expérimentation sociale » privilégiant successivement tels ou tels volets d'action

Pour conclure, « 5.3 Le CESE recommande qu'en en prenant, pour ne pas créer d'illusions, le temps et les moyens nécessaires, l'Union s'engage explicitement à promouvoir, accompagner et soutenir plus activement encore, des expérimentations sociales innovantes dans les champs nombreux, subtils et essentiels des politiques d'inclusion. »

Ces rencontres ont été préparées par une étude du think tank « Notre Europe », coordonnée par Marjorie Jouen, et ayant donné lieu à la production du rapport « *Les expérimentations sociales en Europe : Vers une palette plus complète et efficace de l'action communautaire en faveur de l'innovation sociale.* »⁶⁸¹.

⁶⁸⁰ COMITE ECONOMIQUE ET SOCIAL EUROPEEN, *Avis du Comité économique et social européen sur « Comment l'expérimentation sociale peut elle servir en Europe à l'élaboration des politiques publiques d'inclusion active »*, (2009) (2009/C 100/13).

⁶⁸¹ Marjorie JOUEN, *Les expérimentations sociales en Europe, Vers une palette plus complète et efficace de l'action communautaire en faveur de l'innovation sociale*, Notre Europe, 2008. Accessible sur le site internet www.notre-europe.eu.

Ce rapport relie explicitement la conduite de programmes innovants avec la méthode d'évaluation expérimentale par assignation aléatoire, en précisant que « *pour pouvoir apprécier les résultats de l'expérimentation et comparer à tout moment une population ou une zone traitée avec celle qui ne l'aura pas été - groupe de traitement et groupe de contrôle -, les experts préconisent de procéder par échantillonnage aléatoire, par analogie avec la recherche biologique.* » et revient sur les manières de rendre cette modalité plus acceptable⁶⁸². Cette méthode doit éviter à l'expérimentation réussie de rester sans suite, en facilitant sa capacité à passer à une échelle supérieure, et en convaincant les responsables politiques⁶⁸³. En se référant à plusieurs reprises aux échanges du précédent colloque de la Dares présenté plus haut, l'auteure la présente comme une méthode « *considérée comme étant la plus probante par les chercheurs* », qui produit « *une meilleure « évidence »⁶⁸⁴ des données et, de ce fait, garantit le prolongement de l'expérience ou sa généralisation si elle réussit. Elle coûte bien moins cher que les techniques standards qui supposent la collecte d'une panoplie très large de données. Elle permet d'éviter des controverses et de longues explications pour justifier la méthode d'évaluation retenue.* »⁶⁸⁵.

Lors de ces rencontres, l'expérimentation sociale, définie suivant l'article référent d'E. Grass comme « *une innovation de politique sociale initiée dans un premier temps à petite échelle, compte tenu des incertitudes existantes sur ses effets et mises en œuvre dans des conditions qui permettent d'en évaluer les résultats dans l'optique d'une généralisation si ces effets s'avèrent probants* », est présentée comme un outil de renouvellement de l'action publique. On l'inscrit comme une réponse aux contraintes budgétaires et aux besoins de modernisation des politiques sociales et de « *rationalisation des décisions publiques* », en les fondant sur des « *données scientifiquement prouvées* »⁶⁸⁶. Les intervenants de ces rencontres sont principalement des acteurs politiques ou de hauts fonctionnaires français ou européens. Les chercheurs participant aux débats et ateliers sont exclusivement du J-PAL ou de l'École d'Économie de Paris, on note la présence notamment de François Bourguignon, Philippe Zamora (Crest J-Pal), Francesco Avvisati (EEP-J-Pal). Un atelier intitulé « *Comment évaluer une expérimentation ?* » rappelle que l'« *évaluation peut prendre de multiples formes. Elle repose sur un ensemble de variables, tant quantitatives que qualitatives* ». Néanmoins, elle conclut que « *l'expérimentation dite contrôlée, qui vise à comparer une population au sein d'un territoire où l'expérimentation est mise en œuvre avec une population témoin, est plus exigeante, même si elle est la méthode particulièrement adaptée.* »⁶⁸⁷. Les animateurs de cet atelier témoignent clairement de cette préférence méthodologique, puisqu'on y retrouve en tant que représentants de la communauté scientifique Elizabeth Beasley (Directrice J-Pal Europe), Bruno Crépond (Crest-Insee J-Pal Europe) et William Parienté (J-Pal Europe), tous trois défenseurs de la méthode expérimentale par assignation aléatoire.

Les bases d'une diffusion de l'expérimentation à l'échelle européenne sont lancées, en témoigne la reconnaissance de cet outil par Vladimir Spidla, Commissaire européen à l'Emploi, aux Affaires sociales et à l'Égalité des Chances, en ces termes

⁶⁸² Souligné dans le texte, *Ibid.* p. 29.

⁶⁸³ *Ibid.* p. 3.

⁶⁸⁴ Cet anglicisme, utilisé comme synonyme du mot « preuve », est assez significatif de l'intérêt que ces chercheurs portent à la recherche anglo-saxonne.

⁶⁸⁵ Marjorie JOUEN, *op. cit.* p. 23.

⁶⁸⁶ HAUT COMMISSARIAT AUX SOLIDARITES ACTIVES CONTRE LA PAUVRETE, *Rencontres de l'expérimentation sociale en Europe, Programme*, 21 novembre 2008, p. 5.

⁶⁸⁷ Le terme « contrôlé » fait référence ici à la traduction du terme anglo-saxon *Randomise controlled trial*, que l'on peut traduire par essais contrôlés par assignation aléatoire *Ibid.* p. 13.

« l'expérimentation mise au service de l'innovation sociale peut nous aider à développer des politiques publiques plus efficaces »⁶⁸⁸. Néanmoins, la position sur le couplage entre l'innovation sociale et la méthode d'évaluation expérimentale par assignation aléatoire n'est pas explicite. Concrètement les conclusions majeures de cette rencontre, selon le communiqué de presse du Haut Commissariat, sont⁶⁸⁹ :

- Le besoin d'une coopération renforcée pour promouvoir l'expérimentation sociale qui « si elle est conduite avec suffisamment de rigueur pour assurer son évaluation, profite à tous ». On retrouve ici la notion de bien public mondial développé par Esther Duflo auprès de l'Agence Française de Développement ;
- L'amorce d'un réseau d'excellence sur l'expérimentation sociale, notamment avec l'Ansa, pour partager expériences nationales et européennes, notamment dans le cadre du programme PROGRESS ;
- Un appel à projets communautaires pour promouvoir les expérimentations transnationales, lancé en 2009 et doté 3,5 millions d'euros ;
- Un processus de travail soutenu pour préparer l'année européenne de la pauvreté et de l'exclusion sociale en 2010.

À la fin de l'année 2008, on retrouve ainsi une confirmation de l'émergence d'un *champ d'action stratégique* composé d'acteurs politiques, administratifs et scientifiques unifiés autour d'une problématisation commune présentant l'expérimentation sociale comme l'alliance d'un projet innovant et de son évaluation par la méthode expérimentale par assignation aléatoire, mais aussi comme instrument de renouvellement des politiques publiques répondant à une crise de légitimité. Néanmoins, si les liens entre les membres de ce champ émergent sont de plus en plus fort, ce dernier apparaît encore comme restreint à des *titulaires* et comporte encore peu de *suiveurs* lui permettant des points d'encrage plus stables. Notre analyse de réseau de ce nouveau champ peut nous permettre de préciser les modalités d'évolution de ce champ. Aussi, c'est à cette période que les premiers challengeurs interviennent pour questionner le discours entourant la légitimité de la supériorité de cette méthode comme instrument de *evidence-based policy*.

2.2.2.4. Les premières initiatives de travail institutionnel de déstabilisation

En réponse à la diffusion de ce discours, des initiatives de travail institutionnel de déstabilisation apparaissent. Les premières critiques de cette entreprise sont de natures diverses et émergent de publics différents. Une première critique provient de chercheurs dénonçant la validité des résultats produits par cette méthode, puis de commanditaires déçus, et enfin des sociétés d'évaluation nationales ou européennes qui perçoivent la menace de la présentation de cette méthode comme l'outil de choix de *evidence-based policy*. Ces critiques proviennent donc de champs stratégiques interreliés au nouveau champ stratégique en construction, lesquels se trouvent en partie déstabilisés par cette nouvelle entité capable de capter de nouvelles ressources et qui fournit un travail cognitif discréditant les autres méthodes d'évaluation.

⁶⁸⁸ LE HAUT COMMISSAIRE AUX SOLIDARITES ACTIVES CONTRE LA PAUVRETE, *Communiqué de presse, Bilan des Rencontres de l'expérimentation sociale en Europe*, Présidence française du Conseil de l'Union européenne 2008, 22 novembre 2008.

⁶⁸⁹ LE HAUT COMMISSAIRE AUX SOLIDARITES ACTIVES CONTRE LA PAUVRETE, *Communiqué de presse, Bilan des Rencontres de l'expérimentation sociale en Europe*, Présidence française du Conseil de l'Union européenne 2008, 22 novembre 2008.

Les premières critiques méthodologiques apparaissent avec l'article de Guillaume Allègre à l'été 2008. Cet économiste de l'OFCE (Observatoire Français des Conjonctures Economiques), issu de Sciences Po Paris, revient sur les expériences américaines de l'emploi de la méthode aléatoire, afin de discuter leur validité interne et externe⁶⁹⁰. Si, en théorie, cette méthode peut corriger des biais méthodologiques présents dans les méthodes non expérimentales ou quasi-expérimentales, il démontre que, dans la pratique, elle est exposée à des obstacles menaçant sa validité scientifique⁶⁹¹. Ce travail de déstabilisation du cadre cognitif construit par le champ d'action stratégique est ici particulièrement intéressant, puisque ce chercheur procède à un « détournement » *des alliés* mobilisés par les promoteurs de cette méthode. Il restitue les contextes politiques des exemples d'expériences canadiennes et états-uniennes auxquelles se réfèrent les défenseurs de la méthode aléatoire, pour légitimer son emploi sur le territoire français, mais contrairement à ces derniers, il revient sur les biais méthodologiques présents dans ces expérimentations, et leur instrumentalisation politique. Il contribue ainsi à recadrer les échanges autour du débat américain suscité par ces expérimentations, en soulignant les difficultés d'interprétation des résultats produits par cette méthode dans la pratique. En appliquant ces leçons au RSA, tout en précisant que ce n'est pas une expérimentation aléatoire, il suggère d'être attentif à la baisse des effets positifs des compléments de revenus dans le temps, notamment lorsque l'expérimentation est de courte durée, et souligne que l'échantillon de l'expérimentation n'est pas représentatif de la population potentiellement concernée par la généralisation⁶⁹².

Les critiques de ce discours viennent aussi des commanditaires d'expérimentations. Dans une publication de l'AFD (Agence Française de Développement), traduite en anglais et diffusée aussi dans les réseaux du NONIE (Network of Networks on Impact Evaluation), les commanditaires de l'expérimentation aléatoire sur le micro-crédit au Maroc questionnent l'hégémonie de cette méthode⁶⁹³. Soulignant que l'attention du département évaluation de l'AFD pour les débats autour de l'évaluation d'impact a contribué à « *persuader l'institution qu'une approche standard (un outil de type boîte noire de mesure d'impact applicable à de multiples situations) ne peut pas constituer une réponse satisfaisante à l'insuffisance des connaissances sur les impacts* »⁶⁹⁴. Bien qu'ils aient réuni les meilleurs scientifiques pour appliquer cette méthode et que le porteur soit extrêmement coopératif, dans le cas de l'expérimentation de l'institution de microcrédit Al Amana, les commanditaires relèvent les difficultés rencontrées dans cette expérience. Le nombre limité d'emprunteurs par rapport à ce qui avait été prévu a compromis la mesure de l'impact et a conduit à intensifier la sensibilisation à ce dispositif. Ces difficultés ont augmenté d'un an la période d'exclusion d'accès au

⁶⁹⁰ Comme le rappelle l'auteur de cet article p. 6 « **La validité interne** d'une expérience est sa capacité à produire une estimation non biaisée de l'effet du traitement sur l'échantillon participant à l'expérimentation. Elle est menacée lorsque (1) il existe des différences de caractéristiques entre les groupes de contrôle et de traitement qui ne sont pas aléatoires et qui affectent la réaction au traitement et lorsque (2) le traitement à évaluer n'est pas le seul paramètre qui diffère entre les deux groupes. » p.7 « **La validité externe** d'une expérience est sa capacité à être généralisée à l'ensemble de la population concernée. Pour cela, (1) l'échantillon doit être représentatif de la population, (2) la réaction au traitement et au non-traitement doit être indépendante du nombre de personnes auxquelles il a été donné et (3) la réaction au traitement ne doit pas dépendre du fait que celui-ci ne peut être administré que de façon temporaire lors d'une expérimentation. De plus, pour que l'expérimentation ait un sens, (4) le traitement doit pouvoir être reproduit à une plus grande échelle et (5) l'effet doit être invariant dans le temps. »

⁶⁹¹ Pour accéder à l'intégralité de cette démonstration se référer à cet article accessible sur internet, Guillaume ALLEGRE, *L'expérimentation sociale des incitations financières à l'emploi, questions méthodologiques et leçons des expériences nord-américaines*, Document de travail de l'OFCE n°2008-22, 2008, p. 3.

⁶⁹² Guillaume ALLEGRE, *op. cit.* p. 33.

⁶⁹³ Le Nonie est un réseau composé de DAC Evaluation Network (OCDE), Evaluation Cooperation Group, International Organization for Cooperation in Evaluation et UN Evaluation Group, la publication de ces auteurs est une traduction de celle de l'AFD citée ci-dessous Jean David NAUDET et Jocelyne DELARUE, *Fostering Impact Evaluations at Agence Française de Développement: A Process of In-house Appropriation and Capacity-Building*, Network of Networks on Impact Evaluation, janvier 2008.

⁶⁹⁴ Jean David NAUDET et Jocelyne DELARUE, *Promouvoir les évaluations d'impact à l'Agence Française de Développement : renforcer l'appropriation et l'apprentissage institutionnels*, *op. cit.* p. 6.

microcrédit des membres du groupe témoin et constitué un manque à gagner pour cette institution⁶⁹⁵. Au vu du coût des deux expériences en évaluation aléatoires, qui ont absorbé 25% du budget d'évaluation externe de sa division Evaluation, l'ADF décide alors d'attendre la production de résultats, et leur réception par les décideurs politiques, avant de s'engager dans d'autres expériences de ce type.

La Société européenne d'évaluation, en décembre 2007, publie une déclaration officielle déplorant que la méthode expérimentale aléatoire soit considérée comme la meilleure ou la seule manière rigoureuse et scientifique d'estimer l'impact d'une intervention. Cette déclaration définit le champ de la pertinence de son utilisation ainsi que ses limitations, en rappelant aussi que d'autres moyens d'établir une causalité existent⁶⁹⁶. En France, la Société Française d'Evaluation (SFE) s'intéresse aussi à l'influence grandissante de cette méthode sur le champ de l'évaluation. Déjà, en 2007, est organisée une journée d'étude invitant Marc Gurgand à présenter une communication « expérimentation et évaluation »⁶⁹⁷. Si quelques observateurs, au contact des acteurs de ce nouveau *champ d'action stratégique*, commencent à alerter leurs collègues de l'influence grandissante de cette méthode dans le champ de l'évaluation de l'administration centrale, on ne peut pas encore observer une position commune sur ce phénomène. Le travail institutionnel de déstabilisation reste donc assez marginal pendant cette période, et ne suscite donc pas d'ajustements particuliers du côté des titulaires du nouveau champ d'action stratégique émergent de l'expérimentation aléatoire.

2.2.2.5. Cartographie des acteurs du champ d'action stratégique émergent après son entrée dans les champs stratégiques d'Etat

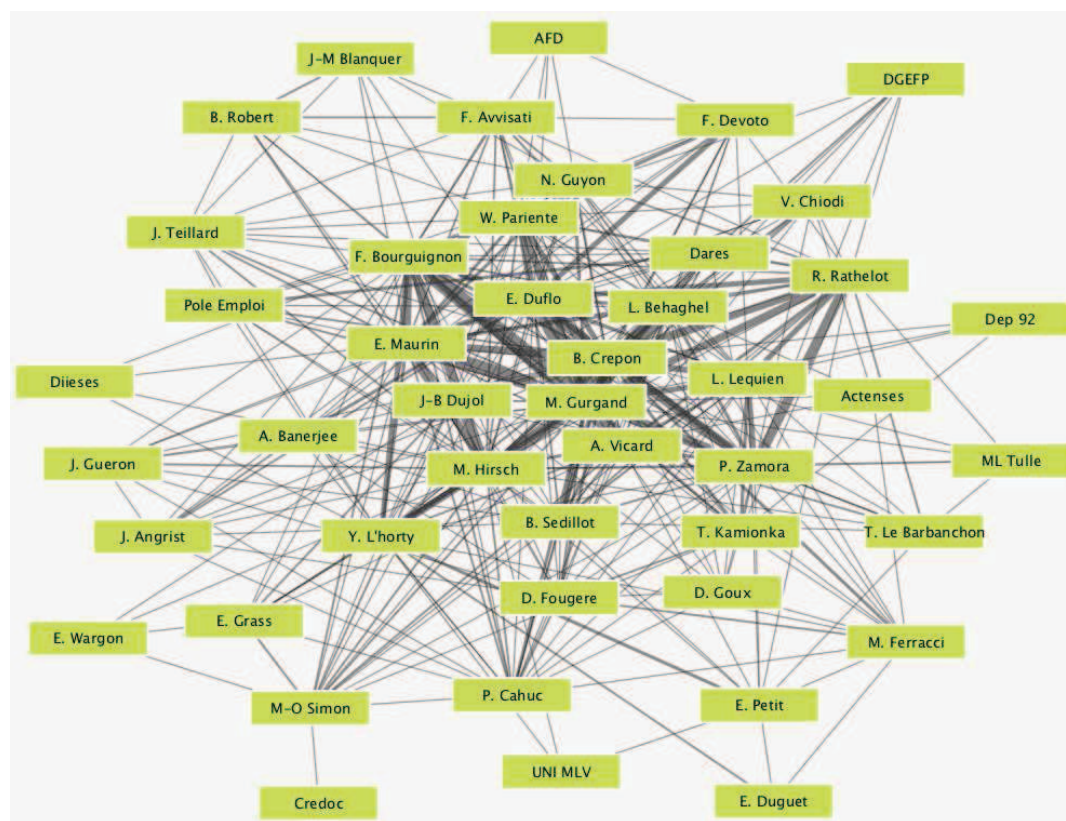
Nous avons aussi pour cette période procédé à une cartographie des acteurs de ce champ d'action stratégique émergent afin de comparer son évolution vis-à-vis de la période précédente. Nous avons rendu visible les mêmes types de liens entre les acteurs, c'est-à-dire : participation à une même communication, scientifique ou institutionnelle, participation à une même expérimentation, publication en commun scientifique ou institutionnelle, et appartenance à la même organisation. Apparaît pendant cette période un champ beaucoup plus étendu comprenant une cinquantaine d'unités. Cette augmentation de la taille du champ est le résultat du processus d'intéressement mené à cette période notamment grâce à la disposition de nouvelles ressources, à la nouvelle connexion de ce champ d'action stratégique aux champs d'action stratégiques d'Etat, et à l'entrée de Martin Hirsch au gouvernement. De nouveaux acteurs de l'action publique sont rentrés dans le champ notamment du fait de l'organisation des premières expérimentations par l'appel à projets du Haut Commissariat aux Solidarités Actives.

⁶⁹⁵ *Ibid.* p. 12.

⁶⁹⁶ EUROPEAN EVALUATION SOCIETY, *EES Statement: The importance of a methodologically diverse approach to impact evaluation-specifically with respect to development aid and development interventions*, décembre 2007.

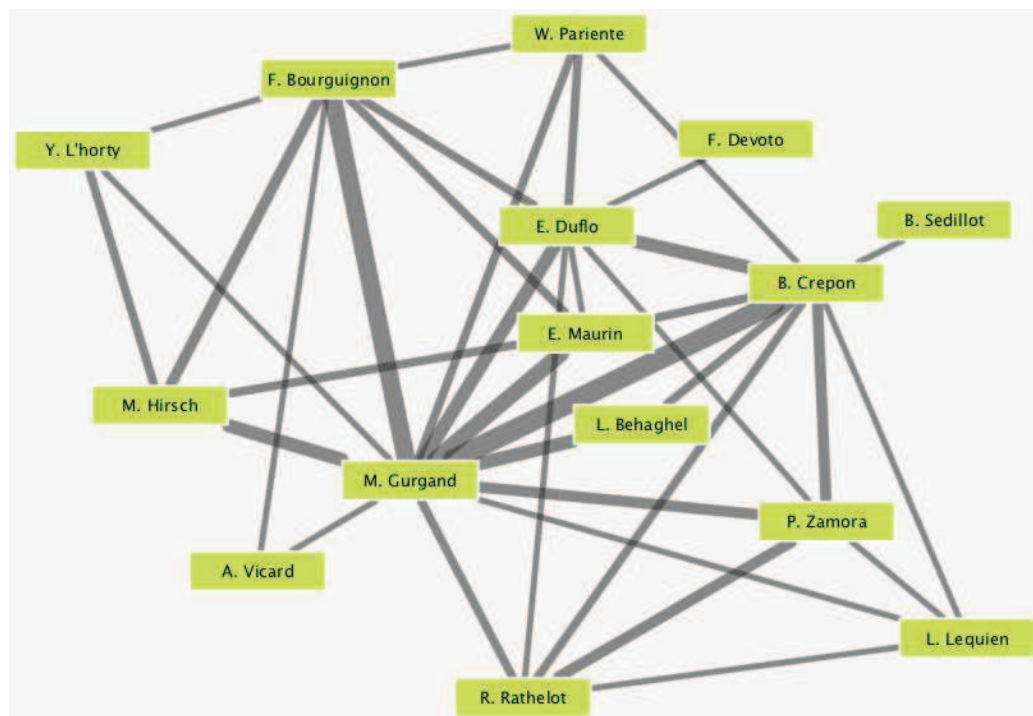
⁶⁹⁷ *Journée d'étude de la S.F.E. L'évaluation et les activités voisines: de la concurrence à la complémentarité*, Programme, 11 décembre 2007.

Figure 9 : Représentation réticulaire du champ d'action stratégique émergent de l'expérimentation par assignation aléatoire de 2007 à janvier 2009 (voir document agrandi en annexe 3)



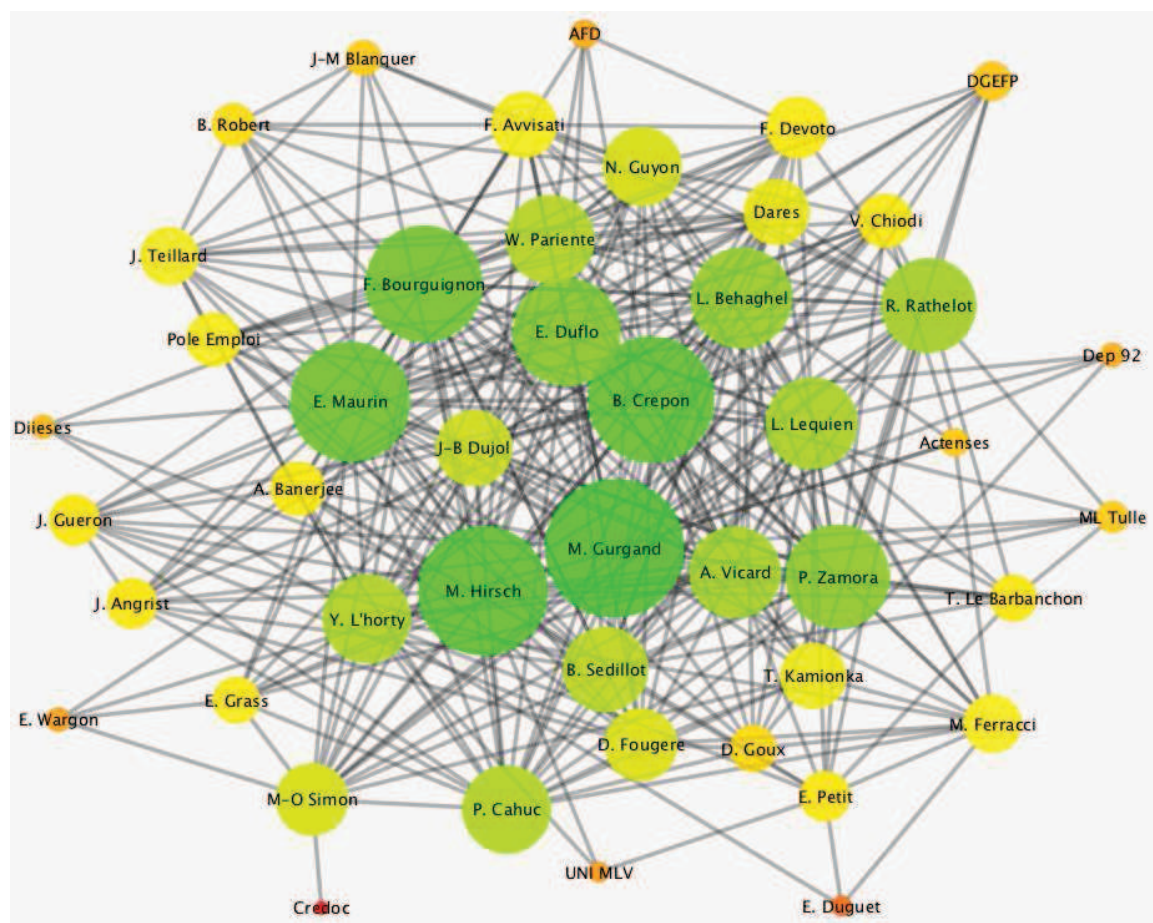
Afin d'identifier le noyau de ce réseau nous avons aussi fait apparaître les individus ayant au moins 4 connections avec le reste des individus du réseau. On observe toujours un noyau dur du champ composé d'acteurs clés reliés entre eux par des liens épais comme B. Crépon, E. Duflo, B. Sédillot de la Dares ainsi que d'autre part M. Gurgand, F. Bourguignon E. Maurin et M. Hirsch aussi très connectés. La Dares et le HCSA restent donc encore pour l'instant les principaux relais de ce champ stratégique au sein de l'Administration Centrale.

Figure 10: Représentation réticulaire du champ d'action stratégique émergent de l'expérimentation par assignation aléatoire de 2007 à janvier 2009, présentant les individus ayant au moins 4 connections avec le reste des individus du réseau (voir document agrandi en annexe 3)



Nous avons aussi pour cette période procédé au calcul du « *average shortest path length* » (longueur moyenne du plus court chemin) et à une mesure du nombre de liens directs que détient chaque individu avec le reste des membres du réseau.

Figure 11: Représentation réticulaire avec calcul de « *average shortest path length* » du champ d'action stratégique émergent de l'expérimentation par assignation aléatoire de 2007 à janvier 2009 (voir document agrandi en annexe 3),



Les individus les plus rapidement reliés à l'ensemble du réseau (en vert) sont encore principalement des économistes du Crest et de l'Ecole d'Economie de Paris et M. Hirsch et la Dares. Quelques nouveaux membres du champ de l'action publique sont entrés dans ce réseau comme B. Robert (rectorat de Créteil), J. Teillard (Cabinet de M. Hirsch), et Pole Emploi, mais ils restent cependant beaucoup moins connectés que les acteurs précédents. En comparant les cartographies d'acteurs de la période 2005-2007 et 2007-2009, on remarque la conservation du même noyau dur mais une intensification des liens entre les acteurs. Si sur la première période on observe un diagramme en forme de papillon, représentant deux pôles, celui-ci prend la forme d'un cercle où les individus sont beaucoup plus connectés entre eux et particulièrement les chercheurs de différents laboratoires. Nous n'avons pu néanmoins continuer cette représentation graphique des acteurs sur les périodes suivantes du fait de l'augmentation du nombre d'acteurs environnant le champ et de la difficulté à caractériser pour chaque individu son lien avec l'ensemble des membres du réseau.

Ainsi, la période de 2007 à 2009 présentée dans ce développement témoigne du travail institutionnel afin d'établir de manière stable un champ d'action stratégique de l'expérimentation sociale. N. Fligstein et D. McAdam, soulignaient dans leur théorie générale des champs la nécessité de connecter un champ émergent avec les champs stratégiques de l'État, afin d'en garantir la stabilité⁶⁹⁸. On observe ici distinctement cette

⁶⁹⁸ Neil FLIGSTEIN et Doug MCADAM, *A theory of fields*, op. cit. p. 77.

stratégie amorcée par le cabinet de Martin Hirsch, engagé dans un travail institutionnel cognitif d'acclimatation de l'administration à la démarche d'expérimentations sociales. Cette démarche devient la ressource principale de ce cabinet pour légitimer son intervention, au sein d'un gouvernement d'une couleur politique idéologiquement différente de la plupart des membres du cabinet, et il ne peut s'inscrire dans les champs traditionnels d'action de l'État du fait de son absence de rattachement à un ministère en particulier. Martin Hirsch, en tant qu'entrepreneur frontière, pratique un difficile équilibre en parvenant, d'une part, à diffuser l'idée d'expérimentations sociales auprès de l'administration et des acteurs politiques, mais, d'autre part, en faisant des concessions sur la méthode scientifique et en mettant en danger l'alliance avec les acteurs scientifiques, comme nous l'avons décrit pour le RSA. Il sait ajuster son discours auprès des scientifiques en s'engageant à promouvoir la commande d'expérimentations sociales évaluées par la méthode expérimentale aléatoire en France et en Europe.

Le contexte français d'internalisation de l'expertise d'État au sein des départements d'étude des ministères, comme par exemple la Dares, offre un terreau favorable au développement de cette méthode qui s'inscrit dans une culture scientifique commune entre les personnes de ces services et les chercheurs promouvant la méthode expérimentale, les deux groupes étant issus de la même école de formation statistique, l'Ensaie et les individus ayant souvent eu des parcours d'aller-retour entre le monde de la recherche et celui de l'expertise interne. En outre, les liens d'Esther Duflo avec l'Ecole d'économie de Paris nouvellement formée favorisent l'installation du J-pal Europe à Paris, structure dédiée intégralement au travail institutionnel international du développement de cette méthode.

Le champ d'action stratégique commence juste à se structurer. Durant cette période, il est principalement composé de *titulaires* qui l'ont forgé à leur avantage. Avec le premier appel à projets, ainsi que les expérimentations dans le domaine de l'emploi, commencent à poindre des *suiveurs* et des *challengeurs*. Ces derniers ont en commun d'avoir un statut moins privilégié dans l'accession aux ressources et de développer des visions alternatives du champ, sans pour autant mobiliser leurs *compétences sociales* pour déstabiliser le champ. On retrouve deux catégories. Les *suiveurs*, ceux qui, par leurs pratiques, contribuent à maintenir et à construire ce nouveau champ, comme par exemple les statisticiens de la Dares soutenant la démarche expérimentale. Enfin les *challengeurs non compétents* qui, par leurs pratiques ancrées dans l'action publique ou associative traditionnelle, contribuent petit à petit à déstabiliser le champ par leur difficulté à se plier aux exigences de cette méthode. Ceux-ci ont à titre d'exemple participé à l'appel à projets de la Dares en proposant des projets innovants mais sans pour autant proposer un protocole d'évaluation expérimentale ou n'ayant pas pu le mettre en place. Malgré tout, le *champ d'action stratégique* des expérimentations sociales reste encore fragile et nécessite, afin de se stabiliser, une structuration autour d'une *instance de gouvernance interne*, ainsi qu'un accès privilégié aux ressources politiques et financières afin de conduire des expérimentations et intéresser (au sens de M. Callon) un nombre de *suiveurs* suffisants.

3. La tentative d'institutionnalisation du *champ d'action stratégique* de l'expérimentation sociale: stabilisation et dissolution

Dans le sillon de la généralisation du RSA, Martin Hirsch ouvre « *un deuxième front* », consacré aux politiques de la jeunesse. Un observateur interne se souvient « *il restait de l'argent sur les expérimentations précédentes et l'on n'avait pas pu travailler sur les moins de 25 ans avec le RSA et notamment sur les questions des ressources des jeunes. Donc il a dealé avec le président de la république le fait de garder ça. Il y a eu un petit article dans la réforme RSA pour dire qu'il était créé un Fonds d'Expérimentation pour la jeunesse, comme ça on sanctuarise, ils étaient très futés Martin Hirsch et son entourage en disant comme ça c'est dans la loi on a un vecteur législatif on sanctuarise l'argent et on a une base de travail* ». L'article 25 de la loi généralisant le RSA, du 1^{er} décembre 2008 prévoit la création d'un fonds d'expérimentations pour la jeunesse (FEJ)⁶⁹⁹. Il est alors nommé Haut Commissaire à la jeunesse le 12 janvier 2009, en plus de son statut de Haut Commissaire aux solidarités actives et reste délégué auprès du premier ministre⁷⁰⁰. Dans son discours « Agir pour la jeunesse », le Président de la République revient sur le cadre et la logique de cette nouvelle mission. Il explique que « *Martin Hirsch va expérimenter* » tout en menant des dispositifs inscrits dans le cadre d'une logique de contractualisation de droits et de devoirs⁷⁰¹. Il présente cette démarche comme nouvelle et ambitieuse, apolitique, puisque soutenue par une personnalité de gauche (Martin Hirsch) orientée vers le résultat⁷⁰². Ces expérimentations sont un processus permettant « *d'attribuer des moyens à la hauteur des résultats obtenus* » et de « *généraliser pour celles ensuite qui fonctionneront* », elles permettent aussi d'étayer le débat politique en témoigne l'évocation par le Président de l'expérimentation du revenu contractualisé d'autonomie : « *Alors je sais que cette idée ne fait pas consensus, mais ce n'est pas une raison pour qu'on la mette de côté et que l'on n'en parle pas. Moi, je pense qu'il faut faire des expérimentations et tester cela avec les garde-fous nécessaires, pour que les termes du débat soient clarifiés* »⁷⁰³.

L'État élève sa contribution à 150 millions d'euros sur 2 ans, la société Total à hauteur de 50 millions d'euros, 3 millions d'euros de l'Union des Industries et des Métiers de la Métallurgie et 2 millions d'euros des contributions d'entreprises ultra-marines au titre de la Lodeom⁷⁰⁴. Ce fond bénéficie de 50 millions de plus fournis par l'État dans le cadre du Plan de relance, pour atteindre un budget de 255 millions d'euros au final. La loi du 27 mai 2009 sur le développement économique des Outre-mer ouvre aussi le financement de projets expérimentaux en direction des territoires d'outre-mer par un abattement d'impôt pour les

⁶⁹⁹ Article 25 « Il est créé un fonds d'appui aux expérimentations en faveur des jeunes. Ce fonds est doté de contributions de l'État et de toute personne morale de droit public ou privé qui s'associent pour définir, financer et piloter un ou plusieurs programmes expérimentaux visant à favoriser la réussite scolaire des élèves et améliorer l'insertion sociale et professionnelle des jeunes de moins de vingt-cinq ans ». LOI n° 2008-1249 du 1er décembre 2008 généralisant le revenu de solidarité active et réformant les politiques d'insertion.

⁷⁰⁰ Décret du 12 janvier 2009 relatif à la composition du Gouvernement, 688 (2009), JORF n°0010. Nicolas SARKOZY, Décret n° 2009-57 du 16 janvier 2009 relatif aux attributions déléguées au Haut-Commissaire à la jeunesse, 994 (2009), JORF n°0014.

⁷⁰¹ « qu'on me comprenne bien, il ne s'agit pas de mieux traiter un jeune qui décroche qu'un jeune, qui peut être en situation sociale difficile, qui suit ses études. Mais le jeune qui décroche, qui se battra, qui assumera ses devoirs, son référent bénéficiera des moyens pour l'aider » Nicolas SARKOZY, Agir pour la jeunesse, Discours de M. Le Président de la République Française, Avignon, 29 septembre 2009, p. 4.

⁷⁰² « (...) je crois que cette politique-là elle n'est ni de gauche ni de droite, elle est d'intérêt général. Et c'est cela qui compte, je sais parfaitement que nous serons jugés, je serai jugé au résultat, j'y suis prêt » Ibid. p. 10.

⁷⁰³ Ibid. p. 8.

⁷⁰⁴ Arnaud RIEGERT, Rapport de stage effectué auprès du Haut-commissaire à la jeunesse du 4 mai 2009 au 21 août 2009, Paris School of Economics, Ecole normale supérieure, Ecole des Hautes Etudes en Sciences Sociales, 2009, p. 10. Programme, Conférence nationale de l'expérimentation sociale « Réformes contre placebo? », Théâtre National de Chaillot-Paris, 29 mars 2010.

entreprises d'outre-mer versant une contribution au FEJ⁷⁰⁵. Ces fonds, gérés par la Caisse des Dépôts et Consignations, sont pluriannuels et sanctuarisés, donc non soumis aux lois annuelles de finance. L'accès aux ressources est donc ainsi sécurisé pour le *champ d'action stratégique* de l'expérimentation sociale, grâce à un soutien de l'État, sans pour autant s'inscrire dans un champ stratégique étatique d'un ministère en particulier.

Ce fonds constitue l'incarnation souhaitée par le *champ d'action stratégique* ayant pour mission d'organiser et de financer la conduite d'expérimentations par assignation aléatoire. Il est conçu comme l'organe français générateur de *evidence-based policy*, par l'organisation de la rencontre du monde scientifique et de l'action publique autour de l'expérimentation. Ce modèle non explicité a été affirmé par M. Hirsch et ses collaborateurs lors de nos entretiens :

(Chercheuse) – Au moment de la création du FEJ, est-ce que vous aviez un modèle étranger en tête ?

(M. Hirsch) – Un modèle de fond ?

(Chercheuse) – Oui.

(M. Hirsch) – Les Etats-Unis. Moi j'ai toujours eu le MDRC parce que c'est ce que je connaissais le mieux. Ceux que je suis allé voir, et à la fois sur leurs thématiques, leur approche, ils me semblaient les plus proches oui.

(Chercheuse) – Est-ce que vous étiez aussi inspirés par les modèles britanniques de l'evidence-based policy ?

H – Oui, un peu, d'abord ce qu'ils avaient fait sur l'éducation, des choses comme ça oui. »

Un de ses membres de cabinet renchérit à propos de l'evidence-based policy : « *Oui absolument c'est une vraie connexion, souvent je disais c'est le moment britannique des politiques sociales françaises, on y est allé on a tout copié, (...) c'est vraiment une connexion importante (...) on était intellectuellement excité par ce qu'il se passait la bas à tel point qu'on y est allé on a fait des visites pour voir comment ça marchait* ».

Il semblerait que les promoteurs de cette méthode, en suivant l'exemple de l'imposition de la norme des essais contrôlés par l'intervention de l'État dans le domaine de la médecine, ont tenté la même entreprise pour les politiques de la jeunesse⁷⁰⁶. Néanmoins, cette institution nouvelle est le théâtre des affrontements entre un travail institutionnel de maintien d'une part et de déstabilisation d'autre part. Incarnation d'un *champ d'action stratégique* hybride composé d'acteurs scientifiques et d'agents de l'État, porté par un entrepreneur frontière, il ne survivra pas au départ de ce dernier et à son inscription de fait dans une relation de subordination hiérarchique aux champs stratégiques de l'État incarnés par les ministères successifs de rattachement du Fonds d'Expérimentation pour la Jeunesse.

3.1. Le Fonds d'Expérimentation pour la Jeunesse comme tentative d'institutionnalisation du champ d'action stratégique

La constitution de ce fonds et l'accession aux compétences de la jeunesse vise à solidifier et institutionnaliser les liens entre les membres du *champ d'action stratégique*, ainsi que d'en étendre l'influence. Ce fonds organise une interaction entre l'action publique et la science autour de la conduite d'expérimentations sociales. La place centrale allouée au conseil

⁷⁰⁵ « *Au 31 décembre 2011, ce sont plus de 5 000 entreprises qui ont contribué à hauteur de 5,5 M€* » FOND D'EXPERIMENTATION POUR LA JEUNESSE, *Rapport d'activité 2009-2011*, 2012, p. 8.

⁷⁰⁶ « *la généralisation de ce type de procédure n'a été possible que grâce à l'intervention de l'État qui imposa la norme des essais contrôlés et randomisés dans les évaluations administratives des médicaments, aux Etats-Unis dès les années 1960, en France, vingt ans plus tard* » Jean-Paul Gaudillière *La médecine et les sciences*, Paris, La Découverte, p90 cité par Agnès LABROUSSE, « Nouvelle économie du développement et essais cliniques randomisés : une mise en perspective d'un outil de preuve et de gouvernement », (2010) 7 *Revue de la régulation*, p.19.

scientifique dans le fonctionnement de ce fonds témoigne de la volonté d'instaurer des critères scientifiques dans le choix de financement de projet. Critères scientifiques dont l'étalon-or est la méthode expérimentale par assignation aléatoire. Cette nouvelle instance permet de sécuriser les ressources dédiées à l'expérimentation sociale, tout en ayant pour souci d'instaurer une instance de gouvernance composée du conseil de gestion qui a pour mission d'assurer le respect de l'expérimentation sociale comme définie par les *titulaires* du champ.

3.1.1. La structuration de l'interaction entre science et politique autour de l'unité de gouvernance interne du Fonds d'Expérimentation pour la Jeunesse

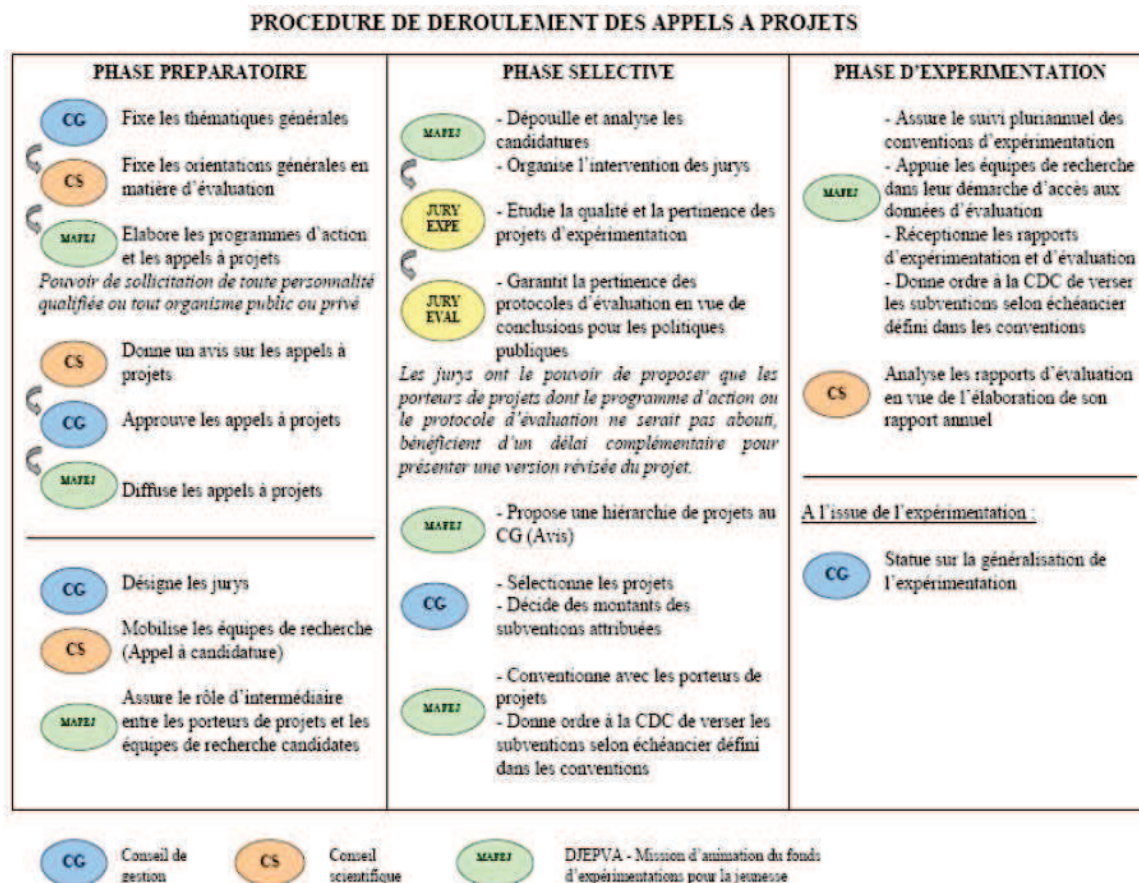
Les instances de gestion de ce fonds organisent la rencontre entre les acteurs de l'action publique et les scientifiques de manière formalisée autour de la sélection de projets. La convention du 22 avril 2009 entre l'État et la Caisse des Dépôts et consignations définit les organes de ce fonds ainsi que leurs fonctions :

- Un organe politique : Le conseil de gestion est présidé par le Haut Commissaire à la jeunesse, il regroupe des représentants des administrations de l'État partenaires et les partenaires privés du fonds, il est responsable de la délibération et de l'administration du fonds, il est présidé par le Cabinet du ministre de rattachement ;
- Un organe scientifique : le conseil scientifique qui fixe les orientations générales en matière d'évaluation pour les appels à projets, « recommande une méthodologie d'évaluation », il peut proposer des sujets d'expérimentations et émet un avis scientifique sur les projets du fonds⁷⁰⁷ ;
- Un organe administratif : la mission d'animation du Fonds (Mafej), une unité de la DJEPVA (Direction de la Jeunesse, de l'Education Populaire et de la Vie Associative), interlocutrice des porteurs de projets qui a des contacts très fréquents avec le cabinet de Martin Hirsch ;
- Un organe de financement : La Caisse des dépôts et consignations est l'intermédiaire financier chargé des versements aux porteurs de projets.

L'interaction de ces instances est permanente dans la conduite des appels à projets. Le conseil de gestion, le conseil scientifique et la Mafej interviennent tous les trois en amont, pendant et en aval des appels à projet. Quelques modalités de sélection ont été modifiées au cours des appels sur lesquelles ce travail reviendra plus tard, mais la structure générale est restée à l'identique.

⁷⁰⁷ Arnaud RIEGERT, *Rapport de stage effectué auprès du Haut-commissaire à la jeunesse du 4 mai 2009 au 21 août 2009*, Paris School of Economics, Ecole normale supérieure, Ecole des Hautes Etudes en Sciences Sociales, 2009, p. 10.

Figure 12: Procédure de déroulement des appels à Projets au 1er semestre 2010



Source : Noblecourt et Valletoux, *L'expérimentation sociale locale en France: bilan, enjeux, perspectives*, p85

Ce cadrage très précis des activités de sélection des projets doit garantir leur correspondance avec le modèle de l'expérimentation sociale promu jusqu'à présent par le champ d'action stratégique. Il correspond à l'unité de gouvernance interne décrite par N. Flitgein et D. McAdam dans leur *théorie générale des champs*. Cette instance de régulation a pour objectif de stabiliser le champ d'action stratégique émergent de manière contrôlée, à la fois en l'ancrant auprès de l'administration via le conseil de gestion, tout en organisant son contrôle par le conseil scientifique composé de chercheurs majoritairement *titulaires* du *champ d'action stratégique*.

Les thématiques générales des appels à projets sont fixées par le conseil de gestion, mais aussi, à partir du second appel à projet, sur la base du livre vert publié le 7 juillet 2009 et intitulé « reconnaître la valeur de la jeunesse » qui fut élaboré par une concertation lancée par le Haut Commissaire le 9 mars 2009.

Composition du Conseil de Gestion interministériel du FEJ

Le conseil de gestion est présidé par le ministre chargé de la jeunesse ou son représentant.
1^{er} collège des représentants de l'État :

- Le directeur de la jeunesse, de l'éducation populaire et de la vie associative ou son représentant ;
- le directeur général de l'enseignement scolaire ou son représentant ;
- le directeur du budget ou son représentant ;
- le directeur général pour l'enseignement supérieur et l'insertion professionnelle ou son représentant ;
- le délégué général à l'emploi et à la formation professionnelle ou son représentant ;
- le délégué général à l'outre-mer ou son représentant ;
- le délégué à l'information et à l'orientation ou son représentant ;
- le directeur général de la cohésion sociale ou son représentant ;
- le directeur des affaires financières des ministères sociaux ou son représentant.

2^e collège des contributeurs privés ou publics autres que l'État.

- la Fondation Total représentée par son directeur général ou son représentant
- l'Union des Industries et des Métiers de la Métallurgie représenté par son directeur général ou son représentant

Le directeur général de la Caisse des Dépôts ou son représentant est invité à toutes les réunions du conseil de gestion en sa qualité de gestionnaire du fonds. Celle-ci dispose d'une voix consultative sauf dans le cas où elle est représentée au conseil en qualité de membre du collège des contributeurs privés ou publics autres que l'État.

Ce conseil de gestion s'est réuni physiquement 5 fois en 2009, 4 fois en 2010 et 5 fois en 2011. Il a été amené aussi à se prononcer par voie électronique sur des points ponctuels et limités entre deux réunions

L'objectif de ce fonds est de faire émerger, de renouveler les politiques de la jeunesse sur la base de la preuve de leur efficacité sur le terrain. Le conseil scientifique est la pièce centrale de cette instance afin de garantir la crédibilité des évaluations conduisant à l'effectivité des projets financés.

Néanmoins, si les procédures de préparation des appels à projets et de sélection des projets sont très cadrées autour de l'interaction de la sphère politique et scientifique, les modalités d'organisation de la décision du passage de l'expérimentation à la généralisation sont très peu précisées. Le conseil de gestion est désigné comme responsable, mais aucun protocole précis n'est explicité. Un acteur interne du fonds précise, en effet, qu'il y avait une recherche d'identification rigoureuse des impacts des dispositifs sans pour autant préjuger d'un modèle de prise de décision publique⁷⁰⁹. La volonté est de créer une boîte à idées dans laquelle les décideurs politiques pourront venir piocher pour conduire de nouvelles politiques. Au-delà de l'intérêt pour l'expérimentation sociale, c'est donc la pratique de l'évaluation par assignation aléatoire qu'il est nécessaire de sécuriser pour les membres du *champ d'action stratégique*. En regardant plus en détail la composition et le discours des instances scientifiques opérant dans le cadre de ce fonds, il apparaît que la conduite d'évaluation expérimentale par assignation aléatoire a été l'un des moteurs principaux à l'origine de cette institution.

⁷⁰⁸ Décret n° 2011-1603 du 21 novembre 2011 relatif au fonds d'appui aux expérimentations en faveur des jeunes, (2011) 2011-1603.

⁷⁰⁹ « on peut se dire qu'on fait de l'évaluation d'impact et qu'on cherche à identifier l'impact sur les politiques publiques sans préjuger un modèle de prise de décision publique »

3.1.2. Une unité de gouvernance dont les critères scientifiques favorisent les *titulaires* maîtrisant la ressource « expérimentation aléatoire ».

La création du fonds et de ses instances offre une tribune étatique à la promotion de la méthode expérimentale par assignation aléatoire. On observe une démarche d'institutionnalisation des liens entre les membres du *champ d'action stratégique* présents dans le cabinet, et ceux au sein des instances du FEJ. La promotion de cette méthode est au cœur de la formation de ce fonds⁷¹⁰. M. Hirsch souligne lors d'un entretien : « *c'était ça l'astuce du FEJ, c'est de venir avec des sous en disant « les sous on vous les met à disposition que si on fait une évaluation réelle, rigoureuse et pas comme d'habitude », donc à ce moment-là on avait une monnaie d'échange assez efficace.* »

La gestion de ce fonds s'opère en lien direct avec le nouveau Cabinet. Celui-ci est organisé en deux pôles correspondant à chaque commissariat, lesquels travaillent en commun. Peu de changements sont opérés sur les membres existants, mis à part l'ajout de nouveaux membres pour occuper les nouvelles fonctions ouvertes par les compétences de la jeunesse (voir liste des membres du cabinet en annexe 4). Un observateur interne de ce cabinet évoque une équipe teintée par la culture scientifique de l'expérimentation et de son évaluation : « *Cela s'est traduit par la présence de personnes ayant une formation scientifique au sein du cabinet, mais aussi par un contact quasi permanent avec des universitaires spécialistes des questions d'évaluation des politiques publiques* »⁷¹¹. En effet, on note une perméabilité croissante entre les mondes de l'action publique et de la recherche, en témoigne par exemple l'arrivée d'Augustin Vicard (économiste et statisticien, administrateur de l'Insee, travaillant juste avant sur l'expérimentation du RSA à la Dares, formé à l'ENS Cachan, à l'Ensaë, et à PSE⁷¹²) comme conseiller technique chargé du fonds d'expérimentation. Le directeur adjoint du Cabinet, Jean-Benoît Dujol, suit une formation à la méthode expérimentale par assignation aléatoire au J-Pal, et le Cabinet accueille un stagiaire de l'École d'Économie de Paris. Comme le souligne un des membres du Cabinet, « *je connaissais très bien cette façon de faire là (l'expérimentation aléatoire), au moins de façon théorique, je l'avais jamais mise en pratique mais je connaissais très bien les chercheurs qui la mettaient en œuvre* ». Ainsi, à partir de 2009, le Cabinet des Hauts Commissariats incarne de plus en plus, la rencontre entre administration centrale et économistes appliqués organisée par le *champ d'action stratégique*. Après la « mission » RSA, le Cabinet s'oriente principalement vers la promotion de la conduite de cette méthode d'évaluation, qui est une des raisons d'être de ce fonds⁷¹³. Cette dimension est précisée sur le site du FEJ : « *L'évaluation occupe une place centrale dans la démarche d'expérimentation sociale portée par le FEJ. Cette démarche d'évaluation est particulièrement ambitieuse, et nécessite pour chaque projet une évaluation externe, par des équipes dont la rigueur méthodologique garantit que les enseignements tirés seront fondés au plan scientifique.* »⁷¹⁴.

⁷¹⁰ Un des responsables de l'évaluation de ce Fonds explique lors d'une allocution à l'institut politique de Lyon que « *La volonté première était d'utiliser les méthodes randomisées pour toutes les expérimentations, mais les contextes institutionnels ne l'ont pas permis.* » le lundi 12 décembre 2011(Mathieu Valdenaire, principes et 1ers retours d'expérience science po Lyon

⁷¹¹ Arnaud RIEGERT, *op. cit.* p. 29.

⁷¹² Après cette parenthèse il retourne travailler à l'Insee dans un département d'études et de recherche, puis au ministère des Affaires sociales et de la Santé dans le département de la Dares au bureau redistribution et évaluation. Il rejoint même l'équipe d'évaluation du revenu contractualisé d'économie (lancé par le FEJ) pratiquant la méthode expérimentale.

⁷¹³ Les membres rencontrés autant du cabinet que de la Mafej s'accordent sur cette mission du fonds en témoigne cette réponse face aux critiques formulées par un interlocuteur « a chaque fois dit c'est inadmissible que 20 à 30% du FEJ ai financé de l'évaluation, mais c'est le lieu d'être c'est la raison d'être de ce fonds »

⁷¹⁴ <http://www.jeunes.gouv.fr/ministere-1001/actions/fonds-d-experimentation-pour-la-1038/#top>

Le Conseil Scientifique du fonds d'expérimentation témoigne aussi d'un processus d'institutionnalisation des relations de conseil tissées pendant les années précédentes avec les scientifiques souhaitant conduire des expérimentations. Le conseil scientifique a été institué le 13 mars 2009 avant même la définition du fonctionnement du Fonds et s'est réuni avant le Conseil de Gestion⁷¹⁵. Il n'a pas de moyens de fonctionnement et ses dix membres sont bénévoles. Parmi ses membres originels, on retrouve Marc Gurgand (EEP) qui en tient la présidence, Yannick L'Horty, Philippe Zamora, et Éric Maurin comme représentants du champ d'action stratégiques. La composition du conseil scientifique est restée stable de 2009 à 2012, seuls deux membres sont partis et trois autres les ont remplacés. Depuis la création du FEJ, le conseil scientifique s'est réuni 15 fois pour apporter son expertise

Encadré 11: Composition du Conseil Scientifique du Fonds d'expérimentation pour la jeunesse⁷¹⁶

Nom	Activité	Période
Marc GURGAND	Chercheur CNRS Ecole d'Economie de Paris, Directeur scientifique J-PAL Europe, chercheur associé Crest, Université catholique de Louvain and EUDN	Président, du 13 mars 2009 à juillet 2013
Matthieu ANGOTTI	Directeur général de la fnars depuis janv 2011, direction du crédoc pendant 3 ans dirige le département d'évaluation des politiques sociales (période 2007-2010)	13 mars 2009
Florence LEFRESNE	DEPP, ministère de l'éducation nationale	13 mars 2009
Yannick L'HORTY	Directeur de la fédération de recherche du CNRS "Travail, emploi et politiques publiques" et dirige le Centre d'Etudes Localisées, Economiques et Statistiques sur les Trajectoires d'Emploi (CELESTE), un des groupes de recherche du Centre d'Etudes de l'Emploi	13 mars 2009
Patricia LONCLE	Enseignant-chercheur (science politique) Ecole des hautes études en santé publique	13 mars 2009
Alberto LOPEZ	Administrateur de l'INSEE, Directeur du Centre d'étude pour l'emploi 2009, directeur scientifique du CÉREQ, direction du CÉREQ d'octobre 2009 à Janvier 2010	13 mars 2009
Philippe ZAMORA	Crest, département de la recherche 2007-2011, 2011 Directeur du CEPE (Centre de formation professionnelle continue du GENES)	13 mars 2009
Éric MAURIN	Directeur d'études EHESS, Ecole d'Economie de Paris	13 mars 2009 jusqu'en août 2010
Vincent VANDENBERGHE	Chef du département d'économie, Louvain School of Economics ,	13 mars 2009 jusqu'en août 2010
Hélène GARNER-MOYER	Depuis 2004, Adjointe au chef de la Mission animation de la recherche Dares, Ministère du travail (Docteur en sciences de gestion de l'Université Paris I)	Depuis septembre 2010
Dominique GOUX	Administrateur hors classe, 2006-2009 mission Animation de la recherche (DARES, ministère du travail). 2009 chercheur au laboratoire de sociologie quantitative du CREST	Depuis septembre 2010
Patrick SILLARD	Insee	Depuis septembre 2010

⁷¹⁵ Source entretien avec une personne interne au fonds et CONSEIL SCIENTIFIQUE DU FEJ, *Rapport du Conseil scientifique du Fonds d'expérimentations pour la jeunesse, pour la période mai 2009-décembre 2010*, Ministère de l'éducation nationale de la jeunesse et de la vie associative, 2011, p. 7.

⁷¹⁶ Ses membres ont été nommés intuitu personae et rétroactivement par un arrêté du 12 janvier 2012 *Arrêté portant nomination au conseil scientifique du fonds d'appui aux expérimentations en faveur des jeunes*, 12 janvier 2012.

La position de cette méthode à la tête de la hiérarchie des méthodes d'évaluation transparaît dans le guide à l'attention des évaluateurs répondant aux appels à projets publiés par ce dernier. Ce guide, rédigé principalement par quelques membres du Conseil scientifique, mais approuvé collégialement, propose une « *présentation complète* » des différentes approches de l'évaluation et recommande fortement le recours à l'expérimentation par assignation aléatoire. Il distingue deux types d'évaluations, l'évaluation qualitative et l'évaluation quantitative, parmi lesquelles « *les méthodes les plus fiables sont les méthodes d'expérimentations contrôlées et randomisées* ». Il est précisé plus loin qu'« *il est toujours souhaitable de privilégier une évaluation selon un protocole d'expérience randomisée dès lors que les conditions de l'expérimentation le permettent.*⁷¹⁷ ». Au total, six pages sur les 22 pages de ce guide sont consacrées à la promotion de la méthode expérimentale par assignation aléatoire comme la méthode la plus rigoureuse scientifiquement. On retrouve plus loin : « *Aux yeux des spécialistes les plus exigeants, ces méthodes sont les seules qui permettent de véritablement prouver les effets d'une expérimentation. La qualité de la preuve est, en quelque sorte, équivalente à celle d'un flagrant délit pour une cour de justice, alors que les autres méthodes, qui mobilisent des données non expérimentales, reviennent à accumuler des présomptions de culpabilité sans jamais véritablement administrer la preuve de façon définitive* ». Le recours aux autres méthodes est légitimé par défaut si cette méthode-là n'est pas applicable.

Au moment de la création de ce fonds, les organes mis en place ainsi que la doctrine établie posent les bases institutionnelles et théoriques de la diffusion de la méthode expérimentale conduite dans la rigueur scientifique de la méthode aléatoire. Néanmoins, dans la pratique, il s'avère que la procédure d'appel à projets ne constitue pas un instrument efficace de mobilisation, mais s'avère être difficilement contrôlable par *l'unité de gouvernance interne*.

3.1.3. Une unité de gouvernance interne hybride difficilement contrôlable par les titulaires du *champ d'action stratégique* : le FEJ « une expérimentation en soi »

Nos interlocuteurs sont unanimes sur le caractère expérimental de la démarche du fonds d'expérimentation⁷¹⁸. C'est une action dans un premier temps peu maîtrisée, qui s'élabore dans la précipitation et dont les équipes se développent en retard sur les besoins en personnel⁷¹⁹. La nouvelle unité de gouvernance conçue pour développer la pratique d'évaluation rigoureuse, et particulièrement celle de la méthode expérimentale par assignation aléatoire, apparaît comme une institution fragile dont peu d'acteurs peuvent garantir l'intégrité vis-à-vis de l'intention de sa création. Si les ressources financières sont un moyen efficace de développement du champ, elles échappent en partie au contrôle du *champ d'action stratégique*, notamment par la procédure d'appel à projets mobilisant le Conseil de gestion et des jurys hybrides. Devant le manque de maîtrise de la procédure d'appel à projets, une centralisation nationale de l'évaluation et une sélection des évaluateurs passant principalement par le Conseil Scientifique est

⁷¹⁷ CONSEIL SCIENTIFIQUE DU FEJ, *Guide méthodologique pour l'évaluation des expérimentations sociales à l'intention des porteurs de projets*, 2009, p. 7.

⁷¹⁸ Le rapport d'activité du FEJ de 2011 souligne p 6 que le FEJ « une expérimentation en soi ». Un membre de la Mafej précise « Mais selon moi il n'y a pas eu de modèle auquel il a été fait référence dans l'action pour dire bon voilà on va s'inspirer de ça ; et d'ailleurs c'est une des caractéristiques, c'est qu'on est quand même parti d'une feuille blanche en essayant de créer nos outils au fur et à mesure »

⁷¹⁹ Comme le souligne un de nos interlocuteurs au FEJ « Jusqu'à aujourd'hui on est quatorze personnes ce qui est quand même toujours assez insuffisant pour gérer 152 expérimentations. Mais à l'époque ils géraient 167, enfin le lancement d'appel à projets et la gestion de 167 expérimentations à quatre. »

privilegiée comme modalité d'incarnation du modèle institutionnel d'origine. Le travail institutionnel cognitif, autant à l'intérieur qu'à l'extérieur, devient de plus en plus intense et apparaît comme le dernier recours pour recadrer l'institution autour des intérêts du *champ d'action stratégique*.

3.1.3.1. Les appels à projets, une stratégie d'intéressement peu maîtrisée par les titulaires

Le processus de diffusion de l'appel à projets n'a été que peu maîtrisé par les membres du *champ d'action stratégique* qui souhaitent promouvoir la méthode expérimentale autant pour l'émergence des porteurs de projets que pour celle des évaluateurs.

La mobilisation des porteurs a été tous azimuts, ce qui a contribué à brouiller le discours autour de l'expérimentation. La diffusion s'effectue par divers canaux, principalement par le Conseil de gestion, par les têtes de réseaux participant à la rédaction du livre vert, par celles ayant participé à la commission Famille, vulnérabilité, pauvreté de 2005, et par les voies hiérarchiques des ministères. Elle fait donc appel aux divers champs stratégiques sectoriels de l'administration centrale. Certains ministères, qui n'ont pas apprécié cette sanctuarisation de financement en faveur de Martin Hirsch, concomitante avec une restriction de leur budget, détournent la raison d'être de cet appel à projet, et le présente auprès des porteurs comme une source de financement potentielle venant pallier aux baisses de financement de l'État. En résulte, l'émergence de porteurs de projets ne s'inscrivant pas dans la démarche expérimentale par choix, lesquels n'ont pas une claire compréhension des pré-requis. Un observateur interne explique : « *Quand vous êtes une grande fédération de l'éducation populaire, qu'on vous retire 200 000 euros de l'État, et qu'en même temps on vous dit : « allez voir Martin Hirsch il a du pognon pour ses expérimentations », et qu'en face on vous pose tout un nombre de contraintes nécessaires pour avoir le droit d'espérer la moitié de ce que l'État vous a pris par ailleurs, vous avez un regard sur l'expérimentation qui est un petit peu différent* ». Malgré les textes de diffusion de l'appel à projets, le discours sur l'expérimentation se brouille, menant certains porteurs de projets à penser que ce financement peut devenir pérenne en cas d'efficacité du projet prouvée par l'évaluation. Au final, cet appel à projet est vécu par beaucoup de porteurs comme une aubaine de financement dans un moment de réduction des budgets. Ainsi, le *champ d'action stratégique* de l'expérimentation sociale, qui était voulu par ses *titulaires* comme indépendants des logiques des champs d'action stratégiques de l'État, se retrouve inscrit dans ces logiques par la présence de représentants de ces champs au sein du Conseil de Gestion et des Jurys de sélection, et donc dans l'unité de gouvernance. Les porteurs de projet sont ainsi encore majoritairement inscrits dans la logique de ces autres champs de l'administration centrale et dans leurs recherches de financements.

Mais la principale difficulté est rencontrée dans l'émergence de l'offre d'évaluation qui ne correspond pas à l'offre souhaitée par les titulaires du champ d'action stratégique. Le Conseil scientifique perçoit cet appel à projets comme un procédé d'intéressement pour faire émerger une offre d'évaluation en provenance des laboratoires de recherche. L'intégralité des interlocuteurs rencontrés composant ce *champ d'action stratégique* estime que l'évaluation pratiquée par les cabinets d'évaluation privés ne correspond pas à l'évaluation d'impact souhaitée, puisqu'elle s'intéresse uniquement au processus. Ils estiment qu'elle est au service uniquement de l'administration des projets et ne s'intéresse pas à l'impact sur les bénéficiaires⁷²⁰. Selon un interlocuteur du FEJ, « *l'évaluateur, ce n'est pas un prestataire de*

⁷²⁰ Comme l'explique un de nos interlocuteurs du FEJ « *il y a un marché de l'évaluation qui existe, qui est celui des cabinets, en fait, qui est souvent une évaluation qui n'est pas une évaluation d'impact en vérité, c'est-à-dire qu'ils s'intéressent pas à*

service, ce n'est pas quelqu'un qui répond à un appel d'offres et qui va être au service de quelqu'un pour apporter la démonstration ». Cet appel d'offres a donc la tâche périlleuse de faire émerger une offre issue des milieux scientifiques en évaluation, alors que, comme ce travail l'a démontré précédemment, la France se caractérise par un désintéressement des chercheurs pour cette pratique.

Une première vague de diffusion est plus ou moins maîtrisée, en passant par les membres du Conseil Scientifique qui préviennent leurs laboratoires ou organismes de recherche respectifs de la parution de cet appel. Ceux-ci peuvent d'ailleurs se constituer aussi comme porteur de projets dans le cadre des programmes, sur le décrochage universitaire par exemple. Sur l'intégralité des projets sélectionnés ayant pour évaluateur un laboratoire, environ 1/3 des projets sont portés par des universités⁷²¹. S'ouvre alors un appel à candidatures sur internet, à destination des évaluateurs, disant que des expérimentations allaient être lancées sur un panel de thèmes différents, et que les évaluateurs potentiels devaient se manifester pour qu'ils soient contactés par des porteurs de projets candidats. Seuls quelques laboratoires répondent à cet appel, ceux proches des membres du Conseil scientifique et quelques cabinets. On retrouve surtout des laboratoires d'économie appliquée, ou des structures comme le Crédoc (Centre de Recherche pour l'Etude et l'Observation) ou le Céreq (Centre d'Etudes et de Recherches sur les Qualifications), qui ont vocation à conduire leur activité de recherche en lien avec l'action publique, et sont sous tutelle des ministères d'État. Ces derniers sont fortement incités à participer à cet appel à projets, source de financements dans un contexte de baisse de leur dotation d'État. Nous retrouvons ici la conjoncture de crise dans ce champ des sciences sociales appliquées que nous évoquons dans notre première hypothèse. Bien qu'ils ne fassent pas partie des *titulaires* du *champ d'action stratégique*, ces acteurs appartiennent à une catégorie de *suiveurs* qui s'inscrivent, par leurs pratiques proches, dans un soutien à ce nouveau champ. Ils ne sont donc pas initiateurs de cette nouvelle norme, mais la soutiennent dans un premier temps, puisqu'elle s'inscrit dans leur intérêt.

Cet appel à candidatures n'est pas suivi d'une labellisation par le Conseil scientifique des évaluateurs, mais aide à constituer un répertoire d'évaluateurs intéressés par thématique, mis à disposition des porteurs de projets. Il y a tellement peu de candidats que certaines thématiques, comme « mobilité internationale », ne répertorient qu'un évaluateur. Les membres du Conseil scientifique estiment que cette initiative de leur part « n'a pas très bien fonctionné » et que cette sur-sollicitation de quelques évaluateurs a pu les amener à standardiser des protocoles d'évaluation entre différents projets, les rendant peu pertinents selon les standards du Conseil Scientifique⁷²². Néanmoins, il est précisé que la liste n'est pas impérative, et les porteurs de projets peuvent sélectionner des évaluateurs hors de cette liste.

Mais le flou et la précipitation entourent les démarches d'évaluation des porteurs de cet appel à projets, dont le délai entre sa publication et sa clôture est seulement de deux mois⁷²³. Peu de

l'évaluation du point de vue des bénéficiaires en essayant de mesurer les effets sur les bénéficiaires, c'est souvent une évaluation qui est plus une évaluation de la mise en œuvre, qui a souvent un positionnement qui est un peu différent de celui qu'on attend nous, qui est un positionnement totalement externe et qui s'adresse à la puissance publique, en général souvent les cabinets quand ils évaluent des politiques publiques non seulement ne le font pas forcément du point de vue de l'impact sur les bénéficiaires, mais en plus le font au service de l'administration, de la collectivité locale qui les embauche ou qui les paye pour les conseillers. »

⁷²¹ Exactement, 31,7% selon notre calcul effectué sur la base des 168 projets répertoriés pour l'API sur le site du fonds d'expérimentation pour la jeunesse.

⁷²² CONSEIL SCIENTIFIQUE DU FEJ, *Rapport du Conseil scientifique du Fonds d'expérimentations pour la jeunesse, pour la période mai 2009-décembre 2010*, op. cit. p. 32.

⁷²³ Publication de l'appel à projets le 10 avril 2009, la clôture initialement prévue pour le 31 mai 2009 est prolongée au 15 juin 2009

précisions sont apportées dans l'appel à projets, mis à part que l'impact des projets doit être rigoureusement évalué. Pour toute demande en matière d'évaluation, les porteurs sont renvoyés à la mission d'évaluation ou d'animation du fonds qui « *pourra appuyer les porteurs de projet pour définir avec eux les conditions d'une bonne évaluation* »⁷²⁴. Ils sont aussi renvoyés au guide méthodologique à l'intention des porteurs de projets qui incite clairement à la mobilisation de la méthode expérimentale par assignation aléatoire, mais n'a pas vocation à répertorier les bons praticiens⁷²⁵. Les porteurs ne savent pas qui solliciter comme évaluateur, le FEJ n'a pas les capacités de les aider au cas par cas. Ils les renvoient alors à ce répertoire, ce qui fait que les évaluateurs se trouvent rapidement débordés par les demandes de collaboration. Ceux-ci doivent donc opérer une présélection des projets puisqu'ils ne peuvent répondre à toutes les sollicitations, ils choisissent les projets s'ancrant au maximum dans leurs préoccupations de recherche, et qu'ils estiment les plus viables et évaluables par les méthodes qu'ils pratiquent⁷²⁶.

Au final, 494 projets sont déposés⁷²⁷. La Mafej n'a pas les capacités en ressources humaines pour gérer ce succès. Elle est structurée en quatre pôles composés, à l'époque, d'une personne pour chacun d'eux, un pôle « suivi et animation de projets », un pôle « évaluation » (composé uniquement d'un stagiaire de l'École d'Économie de Paris jusqu'en septembre 2009, puis d'un chercheur à temps plein ensuite), d'un pôle administratif et financier, et d'un secrétariat⁷²⁸. Les instances du FEJ, et particulièrement la Mafej, se retrouvent donc débordées par le nombre de dossiers à étudier. S'organisent des jurys hybrides composés de chercheurs, de membres de l'administration et d'acteurs des champs d'action publique concernés, présidés par des membres du conseil scientifique. La volonté de préserver le modèle institutionnel défendu par le *champ d'action stratégique* se dissout alors dans les autres intérêts des différentes parties prenantes. Un interlocuteur du FEJ exprime ce ressenti : « *du coup on a évidemment entériné un certain nombre d'évolutions qui se sont faites sans qu'on puisse y faire grand-chose* ».

La ligne directrice à l'origine de la formation du fonds par le *champ d'action stratégique* est uniquement garantie par une partie des membres du Conseil scientifique éparpillés dans les différents jurys. Les critères de sélection des projets, notamment au regard de la qualité de leur protocole d'évaluation selon les standards du guide rédigé par le Conseil scientifique, sont difficiles à maintenir⁷²⁹. D'abord, très peu de projets proposés entrent dans les standards scientifiques requis par ce dernier, mais aussi d'autres critères de nature politique appartenant aux logiques des autres *champs d'action stratégique* de l'État entrent en jeu. En effet, le Conseil de gestion ainsi que les membres des jurys ont été le relais d'autres critères de sélection, prenant en compte les enjeux politiques de la sélection des porteurs de projets. Cette influence des enjeux politiques se reflète dans le nombre de projets acceptés, puisque cet appel à projets était

⁷²⁴ FONDS D'EXPERIMENTATION POUR LA JEUNESSE, *Programme national d'expérimentations pour les jeunes, premier appel à projets 6 avril 2009, 7 mai 2009*.

⁷²⁵ CONSEIL SCIENTIFIQUE DU FEJ, *Guide méthodologique pour l'évaluation des expérimentations sociales à l'intention des porteurs de projets*, 2009.

⁷²⁶ Ces interlocuteurs parlent de présélection puisqu'ils expliquent qu'étant donné les délais impartis pour le dépôt de dossiers, le temps d'une sollicitation par un porteur de l'étude du dossier par le potentiel évaluateur s'il se soldait par un refus, condamnait le porteur à ne pas pouvoir déposer son dossier faute de temps pour travailler sur une offre avec un nouvel évaluateur potentiel.

⁷²⁷ 494 dossiers au total source MAFEJ, *API Statistiques sur les projets déposés et les projets retenus*, HCJ, DJEPVA, 15 mars 2010.

⁷²⁸ CONSEIL SCIENTIFIQUE DU FEJ, *Rapport du Conseil scientifique du Fonds d'expérimentations pour la jeunesse, pour la période mai2009-décembre 2010, op. cit.*

⁷²⁹ Un membre de la Mafej explique les difficultés de compréhension entre les standards de sélections conçus par le FEJ et ceux des autres membres des jurys. Pour chaque projet il cherchait à savoir : « *qu'est ce que ça nous dit sur la conduite des politiques publiques et dans quelles mesures ça va venir alimenter ou en tous cas est-ce que ça a un caractère transposable et généralisable et c'est la question fondamentale qu'il faut se poser, et ça a été une lutte de tout mon temps pour l'expliquer aux porteurs de projets et y compris au jury qui disait ha, mais c'est un projet formidable...et je disais c'est un projet formidable, mais l'évaluation elle est comment elle est nulle* »

vu par certains ministères comme la possibilité de pallier aux baisses de financement accordés à quelques organisations⁷³⁰. Les ministères sont en perte de budget avec la RGPP (Révision Générale des Politiques Publiques) et annexent ainsi une partie du Fonds pour garantir des financements qu'ils ne peuvent plus assurer. Cette tendance est d'ailleurs dénoncée comme un mécanisme d'autofinancement par Matthieu Lamarre, élu de l'Essonne⁷³¹.

Le fonds d'expérimentation se retrouve donc dans une position d'intervention interministérielle conflictuelle au sein d'une action de l'État qui fonctionne par champs stratégiques, ayant chacun leur pré carré, mais manquant de financements. En effet, après les avis des jurys, les services déconcentrés de l'État, notamment de la DRJSCS, ont émis un avis sur les projets, et le Conseil de gestion (l'organe politique du FEJ) était l'arbitre final, permettant ainsi de « repêcher » des projets qui, politiquement, ne pouvaient être mis de côté. Tous les dossiers avec un avis favorable ont été retenus, et certains dossiers dont l'avis du jury n'était pas positif ont été « repêchés »⁷³².

Au total, pour ce premier appel à projets, 494 projets ont été présentés, 325 rejetés (66%) et 169 retenus dont 99 avec demande de modifications⁷³³. Parmi les 60% de projets retenus auxquels sont demandés des modifications, ces dernières portent pour la moitié des cas sur le protocole d'évaluation, soit environ une cinquantaine de projets. Ainsi, environ un tiers des projets sélectionnés ont été « repêchés » aux dépens de leur correspondance avec les standards du Conseil Scientifique. Comme l'explique un interlocuteur témoin de ce phénomène, « *on a vite demandé un peu de souplesse (au conseil scientifique) car on ne pouvait pas avoir des trucs mega robustes donnant lieu possiblement à des publications* ». Un observateur interne explique : « *le conseil scientifique (...) en dehors de l'assignation aléatoire, du randomized experiment, pour être précis, il n'y avait pas de salut. Ils voyaient ça [les évaluations déposées] de manière extrêmement méprisante en disant « mais qu'est-ce que c'est que ça ? ». Et puis au final ils se sont tournés vers le conseil de gestion en disant « ça on ne prend pas, ça on ne prend pas », et le conseil de gestion a dit « non mais attendez, vous rigolez ou quoi ? (...) c'est toutes les Régions de France qui sont en train de nous envoyer des projets et vous dites non ? Au motif qu'on ne va pas faire du double aveugle ? Mais on ne peut pas le faire, les Régions ne sont pas prêtes ». Donc rétropédalage du conseil scientifique, en disant « ben oui, si on fait la méthode (...), il nous restera zéro, ou une ou deux petites expérimentations ». Et donc, il a été très gentiment demandé de bien vouloir réviser le cadre méthodologique, de l'assouplir légèrement, de telle façon à ce que l'intérêt des actions publiques présentées dans ce cadre là puisse être reconnu et qu'elles puissent aller de l'avant* ». Un autre participant aux jurys revient sur les pressions politiques émanant des champs stratégiques d'Etat : « *A la fin du printemps au début de l'été, Il y a eu des pressions politiques dans tous les sens, parce qu'il y avait beaucoup d'argent à la clé, le fonds était quand même doté de 120 millions d'euros je veux dire, donc les Régions étaient tout à fait parties prenantes du truc, et étaient vraiment intéressées. Les arbitrages de comment les projets ont été sélectionnés, ça a été des arbitrages politiques* ».

Les couples de porteurs de projets et d'évaluateurs « au rattrapage » avaient alors jusqu'à fin septembre pour proposer des modifications. Mais, comme le souligne un

⁷³⁰ Un témoin de ces jurys précise « *typiquement sur l'API, ça a quand même été une volonté politique, d'avoir une masse de projets suffisante, alors que les avis des jurys étaient assez lourds et suggéraient que la qualité de l'évaluation n'était pas là.* »

⁷³¹ Matthieu LAMARRE, « Fonds d'expérimentation Jeunesse : ou comment l'Etat s'auto-subsidie », *Paperblog*.

⁷³² Arnaud RIEGERT, *op. cit.* p. 12.

⁷³³ Au final 168 projets seront véritablement engagés, MAFEJ, *API Statistiques sur les projets déposés et les projets retenus*, HCJ, DJEPVA, 15 mars 2010.

interlocuteur au FEJ, il y a eu un jeu entre le Conseil de gestion et le Conseil Scientifique sur le caractère impératif des modifications à apporter pour la sélection du projet. Le Conseil de gestion a retenu les projets et demandé d'appliquer les recommandations méthodologiques, mais sans que ce soit un pré-requis au financement⁷³⁴. Cette modalité a été modifiée pour les autres appels à projets avec un double passage du projet en jury pour juger de la pertinence des modifications apportées. La Mafej a dû assurer pendant cette période le travail de rattrapage avec ces projets très éloignés des critères imaginés par le champ d'action stratégique⁷³⁵. Au total, 38 projets ont proposé d'être évalués par la méthode expérimentale par assignation aléatoire, dont 15 sont portés par des établissements d'enseignement supérieur, 12 par des associations, et 11 par des missions locales⁷³⁶. Au Final, 23 évaluations de cet appel à projet conserveront ce protocole et la moitié parviendront à le mener à termes⁷³⁷.

Les membres du Conseil scientifique tirent les leçons de cette première tentative de grande ampleur d'élargissement du champ de la méthode expérimentale par la procédure d'appel à projets. Ils ont buté sur une contrainte de l'offre d'évaluation, et de la maîtrise de cette nouvelle technologie dans le champ de l'évaluation en France. Certains regrettent que ces financements aient contribué à financer des cabinets privés ou d'autres structures ne maîtrisant pas la méthode expérimentale. Environ 40% des projets retenus sont évalués par des cabinets privés, soit 56 projets. On observe une concentration sur quelques cabinets qui étaient répertoriés sur la liste du FEJ, comme Kaléidoscope (10 projets), Copas (8 projets), Pluricité (6 projets), et Amnyos (5 projets).

Cette allocation des ressources a échappé au contrôle des membres du champ d'action stratégique, qui ont dû faire face à la réticence des autres chercheurs à se concevoir comme évaluateurs au service de la puissance publique⁷³⁸. De fait, les laboratoires ayant répondu présents massivement sont ceux habitués à travailler pour des commandes publiques, comme le Céreq et ses centres associés qui ont remporté 26 projets et 14 évaluations pour le Crédoc. Le conseil scientifique déplore la difficulté à mobiliser les autres laboratoires scientifiques qui « *envisagent moins naturellement la posture d'évaluateur.* »⁷³⁹

L'autre difficulté rencontrée, a été aussi le manque de maturité de l'offre d'évaluation dans la maîtrise de la méthode expérimentale. Cette dimension est soulignée dans le premier rapport du Conseil scientifique : « *dans certains cas, cependant, cette démarche méthodologique semblait artificielle ou peu maîtrisée et affichée dans l'espoir de satisfaire le jury. Si, comme nous l'avons indiqué, il s'agit d'un outil extrêmement pertinent dans certains cas, il peut être mal adapté ou mal utilisé* »⁷⁴⁰. Le Conseil scientifique reconnaît avoir « *explicitement encouragé* » cette méthode, mais regrette qu'elle soit perçue parfois comme indispensable. D'ailleurs, suite à cet appel à projets, les membres de la Mafej proposent à la Société

⁷³⁴ « En pratique, les répondants ont été informés des réserves, mais les dossiers n'ont pas été réexaminés avant acceptation finale, si bien que les difficultés identifiées par les jurys apparaissent maintenant, lors de la mise en œuvre » CONSEIL SCIENTIFIQUE DU FEJ, *Rapport du Conseil scientifique du Fonds d'expérimentations pour la jeunesse, pour la période mai2009-décembre 2010*, op. cit. p. 14.

⁷³⁵ Un acteur du FEJ se souvient : « et après pendant l'été alors qu'on était bien crevé et qu'on avait bossé deux mois la nuit le jour, on a commencé à ouvrir les dossiers et se dire « oh mon dieu » »

⁷³⁶ CONSEIL SCIENTIFIQUE DU FEJ, *Rapport du Conseil scientifique du Fonds d'expérimentations pour la jeunesse, pour la période mai2009-décembre 2010*, op. cit. p. 15.

⁷³⁷ FONDS D'EXPERIMENTATION POUR LA JEUNESSE, *Rapport d'activité 2009-2011, Annexes*, 2012.

⁷³⁸ Un acteur du FEJ explique : « Il y a quand même pas mal de chercheurs, je dis ça sans juger, qui considèrent que ce n'est pas leur métier de se retrouver à évaluer un porteur de projets pour donner des conclusions à l'État, leur métier c'est dans une perspective de beaucoup plus long terme de s'intéresser à telle ou telle politique publique comme objet de recherches, de manière beaucoup plus distanciée, et voilà, il y a quand même du coup je pense une difficulté à faire émerger les gens qui ont la compétence méthodologique en tant qu'évaluateur d'expérimentation, puisqu'en tant que chercheur qui s'intéresse aux politiques publiques, c'est une chose, mais les faire passer de ça à « je suis l'évaluateur d'une expérimentation », c'est plus compliqué »

⁷³⁹ CONSEIL SCIENTIFIQUE DU FEJ, *Rapport du Conseil scientifique du Fonds d'expérimentations pour la jeunesse, pour la période mai2009-décembre 2010*, op. cit. p. 32.

⁷⁴⁰ Ibid. p. 15.

française d'évaluation d'organiser des formations à la pratique de la méthode aléatoire⁷⁴¹. Parmi les évaluations proposant le recours à la méthode aléatoire, seules sept sont menées par trois des structures appartenant aux *titulaires* du *champ d'action stratégique*, les autres sont des laboratoires ou des cabinets ayant choisi cette méthode principalement pour augmenter leur chance de succès dans la sélection du projet, mais qui ne mesurent pas encore les implications de ce protocole⁷⁴².

Le premier appel à projets a donc essuyé les plâtres d'une instance de gouvernance s'appuyant sur la sélection d'appels à projets pour inciter à l'utilisation de la méthode aléatoire. Les évaluateurs et les porteurs ayant sélectionné cette méthode, et rencontrant des difficultés à l'appliquer, remettent en cause sa pertinence à la lumière de ses difficultés d'adaptation au terrain. En retour, les membres du *champ d'action stratégique*, et notamment du Conseil Scientifique, rétorquent que ces projets n'étaient pas adaptés au tirage aléatoire, que les évaluateurs n'avaient pas une assez grande maîtrise de cette méthode, ou qu'elle nécessitait un travail préalable « d'acculturation » des acteurs de l'expérimentation⁷⁴³.

D'autres appels à projets, dont l'évaluation n'est pas d'ampleur nationale, sont organisés. Néanmoins, le processus d'appariement entre évaluateur et porteur est de plus en plus contrôlé par le FEJ. Un second appel à projets est lancé en septembre 2009, le Conseil Scientifique précise alors davantage les attendus de l'évaluation pour chaque thématique de l'appel à projets, afin de garantir la bonne utilisation de la méthode⁷⁴⁴, notamment en augmentant les critères d'effectifs des cohortes afin d'établir la puissance statistique des résultats. Au-delà de la défense de la méthode aléatoire, s'opère le passage à la défense de la bonne conduite de la méthode aléatoire. Un membre du FEJ explique cette nouvelle démarche « *donc effectivement on a écrit : « quand c'est possible et à chaque fois que c'est possible, on va vous demander de vous poser la question de comment et en tout cas d'essayer d'avoir au moins un groupe témoin » on a un peu abaissé nos exigences en disant : on randomise quand on a des vrais enjeux et qu'il y a les volumes, si on a pas les grands volumes au moins, on demande d'avoir une réflexion sur le groupe témoin* ».

Aussi, le Conseil Scientifique se repositionne dans la sélection des projets, en faisant en sorte que les projets sélectionnés, « sous réserve de modification de protocole », repassent pour approbation auprès du Conseil Scientifique, afin de vérifier que les modifications requises ont bien été traitées par l'évaluateur. Aussi, des procédures de sélection en deux temps, avec deux jurys de sélection ont été organisées, le premier regardant les projets en tant que tels, le deuxième regardant les modalités d'évaluation. Dans cette architecture-là, le dernier jugement revient au jury scientifique sur la qualité de l'évaluation.

Au fur et à mesure, l'appariement entre porteurs de projets et évaluateurs est de plus en plus structuré. Dans les derniers appels à projets, à partir de l'appel à projet « lutte contre le harcèlement en milieu scolaire » en 2011, les projets et les évaluateurs sont sélectionnés, dans un premier temps, à part l'un de l'autre. Dans un deuxième temps, une liste de présélection est publiée permettant aux deux groupes de s'apparier. Les projets sont alors présentés conjointement et jugés une deuxième fois par le FEJ. Cette procédure permet de contrôler les évaluateurs que choisissent les porteurs de projets, et

⁷⁴¹ Cette formation ne sera pas finalement organisée, faute des délais de prise de décision liés au statut associatif de la SFE qui s'est avéré trop long pour le FEJ. (source, entretien avec l'ancienne présidente de la SFE)

⁷⁴² On retrouve dans cet appel à projet dans les évaluateurs proposant la méthode expérimentale par assignation aléatoire, le Crest pour 3 projets, l'École d'Économie de Paris pour 1 projet, et le TEPP de Yannick l'Horty pour 2 projets.

⁷⁴³ « *il y a des débats et des difficultés de mise en œuvre des expérimentations. C'est que... il y a toute une acculturation quand même qui est importante* » Ces termes employés par un membre du FEJ témoignent de l'étrangeté de cette culture scientifique pour les acteurs de l'action publique.

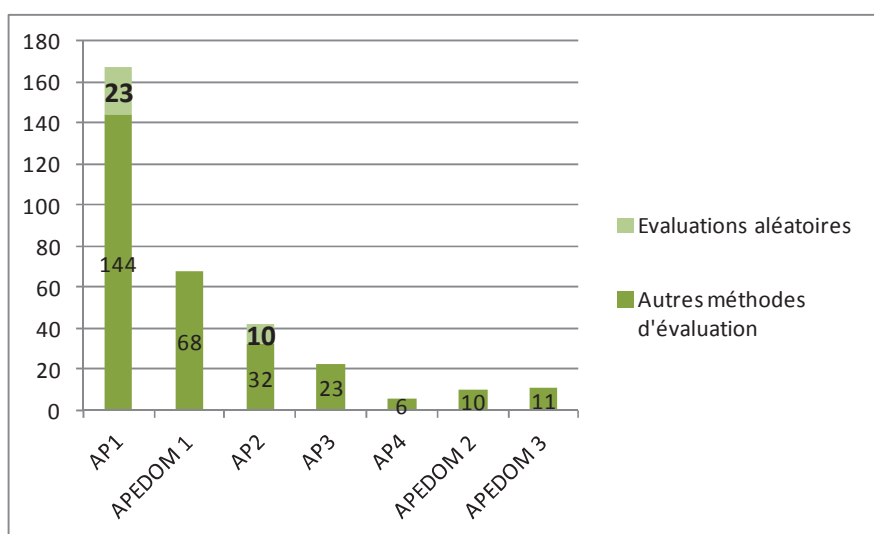
⁷⁴⁴ Un membre du FEJ précise « *Typiquement dans l'AP2 on est à rédigé thématiques par thématiques ce qui était attendu, typiquement pour les discriminations, on a dit voilà, on veut vraiment des évaluations d'impact sur des échantillons larges.* »

de garantir qu'un porteur de projets compétent, ou qu'un évaluateur compétent, ne souffrira pas d'un mauvais appariement. Il permet aussi aux évaluateurs s'investissant dans cette procédure de compétition de prendre moins de risques de passer du temps sur une offre d'évaluation qui ne sera pas sélectionnée. Cette démarche est explicitée dans le rapport d'activité du conseil scientifique en 2011⁷⁴⁵:

« Les porteurs ne sont pas toujours capables d'identifier les évaluateurs qui pourront concevoir les meilleurs dispositifs. Certains évaluateurs proposent des approches « clé en main » qui donnent lieu à des réponses très formatées et pratiquement identiques sur différents projets : les jurys jugent assez mal ce type de réponse, et les porteurs concernés en payent le prix. Dans certains cas, aussi, les porteurs peuvent être tentés de préférer un évaluateur peu exigeant, qui impose peu de contraintes au fonctionnement du projet : cela donne lieu parfois à des évaluations moins solides, ce que les jurys sanctionnent également. Il n'y aurait pas d'autre solution à cela que de constituer des listes d'évaluateurs agréés par le FEJ, mais c'est une sélection qui est très difficile à faire in abstracto. On pourrait éventuellement imaginer de la mener sur la base des rapports d'évaluation à l'issue des premiers appels à projets. »

Au final, la procédure d'appel à projets sans une évaluation nationale centralisée a été peu à peu abandonnée comme outil de promotion de la méthode expérimentale par assignation aléatoire. Seuls l'AP1 (Appel à projets 1) et l'AP2 ont fait émerger quelques utilisations de cette méthode.

Graphique 5: Nombre de projets évalués par la méthode d'évaluation aléatoire parmi les appels à projets dont l'évaluation n'était pas centralisée au niveau national (période 2009-2013)



Source : FONDS D'EXPERIMENTATION POUR LA JEUNESSE, *Rapport d'activité 2009-2011, Annexes*, 2012.

AP1 : Appel à projets n°1

APEDOM 1 : Appel à projets destiné aux départements d'outre mer n°1

AP2 : Appel à projets n°2

AP3 : Appel à projets n°3

AP4 : Appel à projets n°4

APEDOM 2 : Appel à projets destiné aux départements d'outre mer n°2

APEDOM 3 : Appel à projets destiné aux départements d'outre mer n°3

⁷⁴⁵ CONSEIL SCIENTIFIQUE DU FEJ, *Rapport du Conseil Scientifique du Fonds d'expérimentation pour la jeunesse pour l'année 2011*, op. cit. p. 8.

La procédure d'appel à projets s'avère peu efficace et peu maîtrisable comme unité de gouvernance interne pour faire émerger des évaluations correspondantes aux standards du *champ d'action stratégique*. Les projets sont trop restreints pour avoir une puissance statistique, et la qualité de l'évaluation est trop soumise à l'appréciation du porteur de projets. Le FEJ favorise alors des procédures permettant une meilleure maîtrise de la sélection des évaluateurs et des protocoles d'évaluation, notamment en centralisant cette dernière afin de permettre des volumes et des financements plus importants. Ce type de procédure se décline sous la forme d'appels à projets dont l'évaluation est nationale et centralisée, et de procédures de financement de projets hors appels à projets (HAP).

3.1.3.2. L'organisation d'une évaluation centralisée nationalement, un recentrage des ressources du *champ d'action stratégique* sur les titulaires

Après cette tentative d'élargissement du *champ d'action stratégique* par l'appel à projets N°1 peu concluante, le FEJ se replie sur l'organisation d'évaluations maîtrisées par les *titulaires du champ d'action stratégique*. Cette dynamique est rappelée dans le premier rapport du Conseil scientifique : *« le Conseil est conscient que l'encouragement très volontariste donné aux projets de cette nature (expérimentation aléatoire), dans un temps très court, ne va sans doute pas produire que des réussites. De ce point de vue, les évaluations aléatoires d'envergure nationale, ou en tout cas « portées par la puissance publique », sont beaucoup plus probantes, notamment parce que leur échelle permet de mobiliser des effectifs statistiquement pertinents et parce que la construction du protocole et son articulation avec le projet lui-même peut se faire de façon beaucoup plus maîtrisée »*⁷⁴⁶

Le FEJ organise des appels à candidatures nationaux ainsi que des subventions pour expérimentations. Les premiers sont issus de priorités nationales à partir du livre vert ou identifiées par le Cabinet du Haut commissaire, et les secondes sont à l'initiative d'un porteur de projets qui peut proposer une expérimentation de grande ampleur sur des problématiques nationales. Ces deux modalités sont distinguées dans le premier rapport du conseil scientifique, puis certaines sont réunies sous la catégorie « expérimentations hors appel à projets (HAP) », alors que d'autres sont référencées en appel à projets indépendant comme le « livret de compétences » ou « le permis de conduire ». Les appels à candidature procèdent par une définition par la puissance publique de l'expérimentation, puis, dans un deuxième temps, d'un appel aux structures volontaires pour la porter ; ainsi, une expérimentation peut se décliner en une multitude de projets locaux. Ces deux procédures ont en commun d'organiser une évaluation unique et centralisée. L'évaluateur unique est sélectionné directement et de manière indépendante par le Conseil scientifique. À la date du 5 février 2012, cette modalité de contractualisation concernait 225 projets au total, soit 44 % de l'ensemble des projets conventionnés par le FEJ.⁷⁴⁷ On compte parmi ces procédures, l'appel à projets permis de conduire (APPC), le livret de compétences (APLC) ou le revenu contractualisé d'autonomie (APRCA). Cette procédure est privilégiée pour les projets de grande ampleur, aux enjeux nationaux, et soulevant un budget important⁷⁴⁸. Ces grands projets, dont l'évaluation est maîtrisée, sont mis en exergue comme représentant le modèle

⁷⁴⁶ CONSEIL SCIENTIFIQUE DU FEJ, *Rapport du Conseil scientifique du Fonds d'expérimentations pour la jeunesse, pour la période mai2009-décembre 2010*, op. cit. p. 34.

⁷⁴⁷ FONDS D'EXPERIMENTATION POUR LA JEUNESSE, *Rapport d'activité 2009-2011*, 2012, p. 22.

⁷⁴⁸ Un membre du FEJ explique « (...) sur des grosse évaluations qui allaient mobiliser des budgets très importants comme le permis de conduire, là on a envie d'être sûr que l'évaluation sera au rendez-vous puisqu'on ne veut pas, on n'est pas irresponsable, on n'a pas envie de dépenser 30 millions d'euros d'expérimentation, même si ça va à des jeunes, sans être capable d'en tirer des conclusions un peu solides et robustes »

d'expérimentation imaginé par le *champ d'action stratégique*. Un membre du cabinet explique cette stratégie en ces termes : « *Sur ces gros projets là, l'ambition c'était vraiment d'avoir ça (la méthode expérimentale), pour avoir à la fin de quelques années d'existence du fonds jeune, de pouvoir dire, « voilà, au moins sur quelques projets, on a quelque chose de très clair avec des résultats très clairs, et puis après on a un foisonnement d'autres projets qui viennent de la base qui sont intéressants, mais c'est compliqué puisqu'il y en a tellement* ».

Cette procédure dissocie complètement la sélection des porteurs de projets et de l'évaluateur. Pour le recrutement de l'évaluateur, le Conseil scientifique a la maîtrise complète de la rédaction du cahier des charges.

Il peut alors décider de requérir l'emploi de la méthode d'évaluation expérimentale par assignation aléatoire, comme critère de sélection des offres comme ce fut le cas pour l'appel à projet sur les permis de conduire⁷⁴⁹. La sélection de l'évaluateur s'effectue alors dans un entre soi, où les membres du Conseil scientifique et les évaluateurs candidats font partie des mêmes laboratoires, voire sont les mêmes personnes. Par exemple, pour le cas de l'évaluation de l'appel à projets sur les permis de conduire, Yannick L'Horty, se présentant comme évaluateur, a dû se démettre temporairement de sa place au conseil scientifique⁷⁵⁰. Très peu d'évaluateurs répondent à ces offres qui requièrent des compétences très spécialisées, ce sont principalement les laboratoires du champ d'action stratégique et des membres du conseil scientifique comme l'École d'Économie de Paris, ou le TEPP (Fédération de recherche Travail Emploi et Politiques Publiques).

Les travaux d'évaluation peuvent être aussi quasiment internalisés, comme ce fut le cas pour le Revenu Contractualisé d'Autonomie. Cette expérimentation a été préparée par le Cabinet du Haut Commissaire, les administrations concernées, et des chercheurs. Des acteurs du champ d'action stratégique, autant politiques que scientifiques, élaborent conjointement le projet et ses modalités de mise en œuvre et d'évaluation⁷⁵¹.

Cette procédure permet au protocole d'évaluation de ne pas être soumis aux aléas de l'appel à projets, et de conduire la méthode expérimentale dans les règles de l'art⁷⁵². Les difficultés peuvent survenir ensuite, au moment de l'association avec les porteurs de projets, qui doivent contraindre leur action au protocole d'évaluation. Les modalités de cette association délicate, notamment dans le cadre de l'appel à projets, sont traitées dans le chapitre 3 de ce travail.

Sur les 7 expérimentations hors appel à projets amorcées lors de cette période, 6 sont pressenties pour être évaluées par la méthode expérimentale par assignation aléatoire, avec comme évaluateur des chercheurs de l'École d'Économie de Paris⁷⁵³. Sur les 4

⁷⁴⁹ « Le conseil scientifique a insisté pour que cette expérience soit réalisée selon un protocole randomisé, le grand nombre de permis financés étant alors susceptible de permettre la détection d'effets très fins » Arnaud RIEGERT, *Rapport de stage effectué auprès du Haut-commissaire à la jeunesse du 4 mai 2009 au 21 août 2009*, Paris School of Economics, École normale supérieure, École des Hautes Études en Sciences Sociales, 2009, p. 13.

⁷⁵⁰ Il a été remplacé dans ses fonctions par un stagiaire au FEJ et au cabinet venant de l'École d'Économie de Paris Arnaud RIEGERT, *op. cit.* p. 14.

⁷⁵¹ Participent à ces réunions Martin Hirsch, Jean-Benoît Dujol, Augustin Vicard, François Bourguignon, Marc Gurgand, Luc Behaghel « pour atteindre les objectifs de cette expérimentation, il a été établi qu'il fallait chercher dans la mesure du possible à mener une expérience aléatoire à l'échelle individuelle, ce qui génère des résultats d'une puissance bien plus importante à effectif constant qu'une randomisation sur les territoires ». Arnaud RIEGERT, *op. cit.* p. 15.

⁷⁵² Les termes exacts de notre interlocuteur au fonds d'expérimentation pour la jeunesse sont « (...) cela permet de beaucoup mieux contrôler les choses, pour avoir quelque chose de « nickel » au niveau évaluation. »

⁷⁵³ Nous comptons parmi ces expérimentations : l'évaluation de l'internat de Sourdon, Coup de pouce clé, Créa jeunes, Défense 2e chance qui fut abandonnée, la mallette des parents 3ème de Versailles et Groupement de créateurs. Seule l'expérimentation imaginez maintenant ne fait pas appel à l'école d'économie de Paris pour son évaluation

autres grands appels à candidatures élaborés à cette période (certains furent lancés plus tard), deux le sont par la méthode expérimentale : le permis de conduire, évalué par le TEPP, et le revenu contractualisé d'autonomie, évalué conjointement par l'Ecole d'Economie de Paris, le Crédoc et la Dares. Pour ces deux derniers projets, le FEJ engagera plus de 34 millions d'euros, et 1 147 807 euros pour leur évaluation. Au final, selon nos calculs basés sur les rapports d'activité du FEJ, plus de 98% (40/42) des évaluations financées par le FEJ ayant recours à la méthode expérimentale par assignation aléatoire ont été sélectionnées ou - si elles n'étaient pas issues d'appels à projets - ont été élaborées pendant cette période, c'est-à-dire de la fondation de celui-ci au départ de M. Hirsh, au printemps 2010.

Ainsi, il apparaît qu'au fur et à mesure, les promoteurs de la méthode expérimentale par tirage aléatoire ont modifié leur stratégie d'institutionnalisation de cette pratique au sein du FEJ. Ils ont constaté que la procédure d'appel à projets permettait difficilement de contrôler les protocoles d'évaluation, et que ce procédé d'intéressement se confrontait à une émergence difficile de l'offre par de nouveaux laboratoires. Ils ont alors concentré leur travail institutionnel sur des procédés leur accordant une approche contrôlée du protocole d'évaluation pratiqué par les acteurs titulaires du champ d'action stratégique et dont ils maîtrisaient les unités de gouvernance internes. Si les procédés d'intéressement d'autres potentiels praticiens d'évaluation aléatoire ont été peu à peu mis de côté, ceux en direction des porteurs de projets ont été constants, et se sont intensifiés pendant cette période. Cette procédure d'émergence d'expérimentations témoigne de la reconfiguration de l'interaction entre l'économie appliquée et l'action publique amorcée dans le premier appel à projets lancé par le Haut Commissariat aux Solidarités Actives en 2007. Les économistes composant le champ d'action stratégique émergent participent activement à la formulation du projet afin de le rendre évaluable par la méthode expérimentale, mais pour que les résultats de l'expérimentation procurent un apport en connaissances scientifiques permettant de s'inscrire dans les débats de micro-économie d'actualité. Néanmoins, la qualité de cette production de connaissance dépend de l'adhésion des porteurs de projets aux exigences du protocole de l'expérimentation aléatoire. Ce soutien nécessite au préalable un travail cognitif intense d'acculturation à la méthode aléatoire, pour que les porteurs de projets la considèrent comme la méthode la mieux à même de prouver l'efficacité de leur projet, et décident de s'en saisir et de se plier à son protocole.

3.1.3.3. La nécessité d'un travail cognitif en parallèle pour asseoir la position des titulaires

« Si en face il n'y a pas quelqu'un qui est convaincu d'une vraie plus-value à faire avec cette méthode, parce qu'il y a de gros problèmes dans la façon dont on évaluait les politiques publiques avant et de manière traditionnelle, ça n'a aucune chance de marcher ». Ces propos d'un titulaire scientifique du champ d'action stratégique illustrent en partie les raisons pour lesquels ces acteurs ont continué le travail institutionnel cognitif amorcé avant 2009. Bien que des protocoles de méthode aléatoire aient été amorcés, ceux-ci rencontrent beaucoup de difficultés à se faire accepter sur le terrain, surtout par les agents de la mise en œuvre du programme qui doivent ajuster leurs pratiques pour qu'elles correspondent à ce protocole⁷⁵⁴. Pour accepter cette contrainte, il faut que ces acteurs soient convaincus de la supériorité de cette méthode sur les autres méthodes d'évaluation moins contraignantes. Cette opération de persuasion ne peut pas se faire par une allocation de ressources ou par un changement de normes, elle passe par

⁷⁵⁴ Ces nécessités de négociations sont étudiées au chapitre 3

un travail cognitif de diffusion d'un discours sur l'évaluation expérimentale par assignation aléatoire. Celle-ci s'opère en interne au FEJ, mais aussi en partenariat avec des chercheurs extérieurs.

Pour la communication du FEJ, le Haut Commissaire confie à deux élus locaux, M. Valletoux, maire de Fontainebleau (élu UMP), et M. Noblecourt, adjoint au maire de Grenoble et président du CCAS (élu PS), la rédaction d'un rapport sur l'expérimentation sociale pour « *assurer la diffusion la plus large possible de cette pratique afin d'en faire un instrument propre à générer de nouvelles politiques publiques* »⁷⁵⁵. Ces élus sont partis à la rencontre des porteurs de projets et des élus locaux en province dans le but de sensibiliser les acteurs et d'évaluer leurs besoins d'appui dans cette démarche. Ils organisent des réunions locales afin de parler des expérimentations, mais aussi pour faire se rencontrer le monde politique et le monde de la recherche. Selon l'un des rédacteurs de ce rapport, le manque de dialogue entre ces deux mondes est patent. Ce travail s'inscrit dans une dynamique de diffusion de la culture de l'expérimentation auprès des collectivités locales, perçues comme un des terrains féconds pour le développement de l'expérimentation, et pouvant néanmoins constituer un blocage comme cela a été le cas lors du RSA. Cette mission devait aussi donner lieu à la création d'un centre de ressources pour l'expérimentation sociale des collectivités locales pour un appui méthodologique, et pourrait avoir la capacité de proposer la généralisation d'expérimentations par les politiques publiques locales⁷⁵⁶. Ces acteurs politiques ont donc servi, en partie, de promoteurs locaux de l'expérimentation, en sensibilisant les collectivités à cette pratique, mais aussi les chercheurs participant à ces réunions.

Un autre travail institutionnel de taille est amorcé par Esther Duflo, présente sur le territoire français pendant l'année 2009. Au cours de cette année, elle se place au côtés de Martin Hirsch, pour renforcer son crédit scientifique. Elle bénéficie d'une tribune scientifique de choix, via l'invitation de l'Agence française de développement (AFD) à occuper une Chaire annuelle « savoirs contre pauvreté » au collège de France⁷⁵⁷. Cette chaire est financée par l'AFD, et témoigne d'un soutien continu de cette agence à la diffusion de l'évaluation expérimentale par assignation aléatoire depuis 2004⁷⁵⁸. Martin Hirsch souscrit volontiers à cette démarche et s'affiche aux côtés de la chercheuse à cette occasion. La leçon inaugurale d'Esther Duflo reprend le discours de problématisation qu'elle avait présentée à l'AFD en 2004 en insistant sur la place des économistes.

En effet, lors de celle-ci le 8 janvier 2009, elle présente la méthode expérimentale par assignation aléatoire, « quasiment inconnue il y a 10 ans », comme une troisième voix entre l'alternative proposée par des économistes du développement promouvant la résolution de la pauvreté soit comme Jeffrey Sachs, par un apport de fonds conséquents, soit comme William Easterly, par le retrait de fonds laissant place aux mécanismes du marché⁷⁵⁹. Elle présente un programme d'action orienté vers la conduite « d'expérimentations créatives » et légitime la participation des économistes au design de ces expérimentations, du fait de leur connaissance des théories de mécanismes de prise de décision par les agents. Cette intervention, soutient le passage de l'économiste appliqué « d'évaluateur à co-expérimentateur », et légitime ainsi la collaboration entre les chercheurs du Conseil Scientifique du fonds d'expérimentation et le Cabinet dans la

⁷⁵⁵ HAUT COMMISSARIAT A LA JEUNESSE, *La mission Valletoux-Noblecourt*, 2009.

⁷⁵⁶ HAUT COMMISSARIAT A LA JEUNESSE, *La mission Valletoux-Noblecourt*, 2009.

⁷⁵⁷ Esther DUFLO, « Chaire Savoirs contre pauvreté », *La lettre du Collège de France* 2009.25.5.

⁷⁵⁸ Cette chaire annuelle a été créée à l'initiative de P. Kourilsky et P. Rosanvallon.

⁷⁵⁹ La vidéo de l'intégralité de cette leçon est consultable sur le site du collège de France à la page dédiée à cette chair : <http://www.college-de-france.fr/site/esther-duflo/inaugural-lecture-2009-01-08-18h00.htm>

conception d'expérimentations de grande ampleur, qui ont vocation à un double usage d'apports théoriques en économie et de tests d'intervention publique. Elle reprend le même discours de problématisation, discréditant d'une part les acteurs politiques qui ne font aucun bilan, ainsi que les « experts » qui ne disposent pas d'outils pour appréhender les faits de manière indéniable et rigoureuse. Elle présente alors l'évaluation expérimentale par assignation aléatoire comme une nécessité pour répondre à cette crise et résoudre, pas à pas, les problèmes de la pauvreté. Elle fait profiter de cette estrade académique de grande ampleur à Bruno Crepon, entrepreneur clé du champ d'action stratégique en France, en lui offrant une séance de cette chaire pour présenter l'expérimentation de Pôle emploi sur les opérateurs privés de placement.

Bien que ce discours ne s'adresse pas directement aux porteurs de projets du Fonds d'expérimentation pour la jeunesse, il contribue, par le prestige de cette position, à entériner la légitimité scientifique de cette méthode.

Les travaux d'Esther Duflo se diffusent au-delà des canaux de l'économie du développement. Elle publie un article dans la Revue d'Économie politique qui acte le développement fulgurant de cette méthode dans les milieux du développement, et répond aux principales critiques adressées contre cette méthode⁷⁶⁰. Ses publications et communications se multiplient. Dans *laviendesidées.fr* elle explique comment « véritablement tester l'efficacité d'une politique publique »⁷⁶¹.

Elle est présentée aussi dans une campagne de presse favorable. Un article du quotidien *le Monde* la décrit comme une « surdouée » revenant au pays, et ayant une fibre à la fois « humanitaire et sociale »⁷⁶². Ce discours se diffuse aussi dans la presse féminine, avec un entretien dans *Madame Figaro* où elle présente l'activité du J-pal, et où elle est introduite en ces termes : « À 36 ans, cette brillante économiste est la benjamine des professeurs du Collège de France. Son approche de l'aide au développement pour les plus pauvres fait des émules dans le monde entier. Sur le modèle des essais cliniques, elle défend les expériences de terrain, avec un droit à l'erreur. Inédit. »⁷⁶³.

Martin Hirsch s'associe donc volontiers publiquement à cette chercheuse, afin de légitimer sa démarche expérimentale, et de renforcer sa crédibilité scientifique. Ils interviennent ensemble à France Culture dans l'émission « Du grain à moudre » ou dans le *Nouvel Observateur* dans une interview intitulée « comment lutter contre la pauvreté : le face-à-face Esther Duflo-Martin Hirsch », malgré le titre de cette interview, c'est bien une démarche conjointe présentée ici - en janvier 2010⁷⁶⁴. Les effets de ce travail cognitif sont soulignés par le Conseil scientifique dans son rapport marquant la fin de cette période, qui estime que la perception des méthodes aléatoires a évolué, et que leur intérêt est mieux compris notamment depuis que des exemples de réalisation sont disponibles⁷⁶⁵.

Ainsi, s'opère au fur et à mesure un décalage, de plus en plus grand, entre l'activité du FEJ qui, en pratique, conduit un faible nombre d'expérimentations évaluées par la

⁷⁶⁰ Abhijit BANERJEE et Esther DUFLO, « L'approche expérimentale en économie du développement », (2009) 119-5 *Revue d'économie politique* 691- 726.

⁷⁶¹ Esther DUFLO, *L'économie du développement à l'épreuve du terrain*, La vie des idées.fr, 5 mai 2009.

⁷⁶² Catherine SIMON, « Esther Duflo: Des neurones contre la misère », *Le Monde* (01 2009).

⁷⁶³ Les mots surlignés le sont dans le texte d'origine, Esther DUFLO, « Esther Duflo, pratiquer l'économie comme une vraie science humaine », *Madame Figaro* (28 février 2009).

⁷⁶⁴ François ARMANET, « Comment lutter contre la pauvreté: le face-à-face Esther Duflo-Martin Hirsch », *Le Nouvel Observateur* n°2359 (21 janvier 2010), p. 70-73.

⁷⁶⁵ CONSEIL SCIENTIFIQUE DU FEJ, *Rapport du Conseil scientifique du Fonds d'expérimentations pour la jeunesse, pour la période mai2009-décembre 2010*, op. cit. p. 34.

méthode expérimentale par assignation aléatoire, mais dont son discours, légitime toute sa démarche par cette méthode scientifique.

Les critiques commencent à émerger dans le champ scientifique. La plus notable est celle émise par le comité éthique du CNRS (COMETS), qui adopte un avis le 19 janvier 2010 sur « l'éthique de la recherche dans l'expérimentation sociale », et rappelle les frontières entre recherche et décision politique, en soulignant que la confusion des deux est potentiellement présente dans l'expérimentation sociale⁷⁶⁶. Il montre la délicate position politique du chercheur qui, par la méthode aléatoire, se place comme « co-expérimentateur ». Il souligne aussi, le phénomène de « cannibalisation » des financements de la recherche en science sociale par les laboratoires pratiquant cette méthode, et donc le glissement vers des financements des sciences sociales uniquement orientées vers l'étude de l'action publique. Il dénonce aussi les risques d'instrumentalisation des résultats de la recherche produite dans ce cadre, comme pour le RSA. Il émet quatre recommandations : préciser la nature de la contribution du chercheur lorsqu'il participe à une expérimentation aléatoire, ouvrir à la pratique d'autres méthodes de recherche plus « observationnelles », faire en sorte que les laboratoires aient leurs propres fonds pour mener des recherches expérimentales, s'inscrire alors dans les normes d'une recherche scientifique, pour que le chercheur soit responsable des conclusions qu'il produit.

Le 22 mars 2010, Martin Hirsch, ne souhaitant pas s'installer durablement dans ses fonctions d'État, met fin à sa mission au gouvernement en étant nommé le même jour président de l'Agence du Service Civique⁷⁶⁷. Dans un entretien au Monde intitulé « Il y avait un climat dans lequel je ne me sentais pas à l'aise », il explique que ce choix tient au maintien de son indépendance, et rappelle que l'expérimentation était au cœur de sa période au pouvoir puisqu'il la cite comme facteur d'efficacité. Il exprime d'ailleurs son souhait de suivre les expérimentations lancées, si le Président le permet, mais cette faveur ne lui sera pas accordée⁷⁶⁸.

Au total, pendant la période où Martin Hirsch était Haut Commissaire à la jeunesse, en moins d'un an, 291 expérimentations sont conventionnées⁷⁶⁹. La dernière grande conférence, incarnant le champ d'action stratégique entre acteurs de l'action publique et scientifiques, est organisée le 29 mars 2010. Cette « Conférence nationale de l'expérimentation sociale », ayant pour sous-titre « réforme contre placebo ? », propose de présenter des expérimentations par le couple porteur et évaluateur. Les expérimentations évaluées par méthode aléatoire et par le groupe de chercheurs identifié précédemment sont surreprésentées. Sur neuf des projets présentés, sept sont évalués par la méthode expérimentale et mobilisent les scientifiques du Crest, de l'École d'Économie de Paris ou du TEPP⁷⁷⁰. Esther Duflo y participe aussi, elle affirme la filiation avec les essais cliniques et salue la multiplication des « laboratoires vivants »⁷⁷¹. Les mérites de la méthode d'évaluation expérimentale sont mis en avant avec la mobilisation d'alliés. Celle-ci est présentée par P. Zamora (Crest) comme parlant

⁷⁶⁶ COMETS, *Avis du COMETS sur « l'éthique de la recherche dans l'expérimentation sociale »*, 19 janvier 2010.

⁷⁶⁷ *Décret du 22 mars 2010 relatif à la composition du Gouvernement*, 5713 (2010), ORF n°0069.

⁷⁶⁸ Martin HIRSCH, « Martin Hirsch "Il y avait un climat dans lequel je ne me sentais pas à l'aise" », *Le monde* (23 mars 2010).

⁷⁶⁹ Chiffres indiqués dans le livret d'accueil de la « Conférence nationale de l'expérimentation sociale, réforme contre placebo ? » du 29 mars 2010.

⁷⁷⁰ On retrouve Philippe Zamora (Crest), Bruno Crépon (Crest), Elise Huilery (Science-po Paris JPAL EEP), Yannick L'Horty (CEE), William Parienté (UCL Jpal EEP), Marc Gurgand (EEP), Luc Behaghel (EEP), voir le détail dans le programme de la conférence *Programme, Conférence nationale de l'expérimentation sociale « Réformes contre placebo ? »*, Théâtre National de Chaillot-Paris, 29 mars 2010.

⁷⁷¹ Le contenu des interventions n'est pas tiré d'acte, mais des notes d'observation qui ont été prises lors de notre participation à cette Conférence.

simultanément aux décideurs praticiens et scientifiques, tout en constituant un langage commun entre eux. M.-P. Daubresse, qui reprend le fonds sous son ministère de la Jeunesse et des Solidarités actives suite au départ de Martin Hirsch, s'inscrit dans cet héritage et se présente comme « *fan* » de cette méthode des « *groupes tests/groupes laboratoires* », bien que, comme nous allons le voir plus loin, sa conduite effective du fonds s'est distancié fortement de cet héritage. La table ronde finale, « Comment passer des expérimentations aux politiques publiques ? », réunit Esther Duflo, Martin Hirsch et François Bourguignon pour un dernier dialogue. Ils reviennent sur le cas de l'évaluation du RSA, victime de « l'économie politique ». La nécessité de se détacher du temps politique est soulevée en proposant notamment de s'orienter vers des financements privés et d'élaborer des bases de données dans lesquelles les politiques pourraient aller piocher. Martin Hirsch précise une dernière fois que le terme expérimentation doit être réservé à des projets évalués par la méthode expérimentale par assignation aléatoire, ou s'en rapprochant. On observe ici encore, le décalage entre le discours autour de l'expérimentation qui présente l'évaluation aléatoire comme la norme évaluatrice du Fonds d'expérimentation alors qu'elle reste sous-représentée en pratique. Après le départ de cet entrepreneur frontière capable de faire des ponts entre l'action publique et les attentes des scientifiques, ce décalage entre les expérimentations conduites par le fonds et le modèle d'expérimentation par assignation aléatoire à l'origine de sa création ne cesse de se creuser.

3.2. Le détournement d'un champ d'action stratégique par la captation de son unité de gouvernance par les challengeurs

Suite au départ de Martin Hirsch, son cabinet politique est dissous, et seule reste la structure du Fonds d'expérimentation pour la jeunesse avec la Mafej, le Conseil Scientifique et le Conseil de Gestion. Avec la présence de Martin Hirsch, cette *unité de gouvernance* du champ a bénéficié d'une indépendance des champs stratégiques des ministères, grâce à son rattachement au Haut-commissariat à la jeunesse, lui-même rattaché directement au premier ministre. La perte de ce privilège provoque des changements radicaux pour le Fonds d'expérimentation pour la Jeunesse, lequel va être successivement affecté au ministère de la Jeunesse et des Solidarités actives puis au ministère de l'Éducation nationale, de la jeunesse et de la vie associative, et sera inscrit de fait dans leur logique de champ, sous l'autorité du ministre ou du secrétaire d'État compétent. Comme l'explique un de nos interlocuteurs dans cette instance : « *le FEJ n'est pas une agence indépendante, (...) ce n'est pas une autorité administrative indépendante, on est dans le giron d'un ministère et d'ailleurs ça explique des orientations un peu différentes selon les époques* ».

Le FEJ, organisation conçue pour promouvoir la pratique de la méthode expérimentale par assignation aléatoire, se trouve détourné de sa raison d'être originelle. Avec le départ de Martin Hirsch, et des deux Hauts Commissariats ainsi que de leur Cabinet, le champ d'action stratégique perd la quasi intégralité des membres de son relais politique, qui avaient jusque-là promu au sein de l'administration centrale l'expérimentation aléatoire. À partir de 2010, l'importance donnée au financement d'évaluations par le FEJ est remise en cause, il est peu à peu récupéré par les logiques ministérielles. Le Conseil Scientifique, ainsi que quelques membres de la Mafej, se retrouvent isolés comme derniers relais de « l'esprit » du FEJ. Le champ d'action stratégique se dissout peu à peu. Le pendant scientifique de ce dernier reste actif et continue à mener un travail institutionnel cognitif autour de la promotion de la méthode expérimentale, mais de façon moins intense, notamment du fait qu'ils sont déjà occupés à conduire les

expérimentations en cours. Ils cherchent aussi, par ailleurs, d'autres relais pour en mener de nouvelles. Quant au personnel du Cabinet, membres du champ stratégique, ce groupe de hauts fonctionnaires se disperse dans des activités diverses. Néanmoins, bien que la méthode expérimentale ne soit plus centrale dans leur activité, ils l'intègrent toujours si possible dans leurs actions. On retrouve ces hauts fonctionnaires, plutôt de sensibilité socialiste, dans les cabinets ministériels de la nouvelle majorité de gauche en 2012. De plus, les économistes appliqués étant toujours en lien avec les départements d'expertise des ministères continuent à exercer leur influence. Mais cet ensemble d'initiatives ne correspond plus aux relations d'interdépendance et de proximité d'un champ d'action stratégique de l'expérimentation sociale, et l'heure est davantage au bilan qu'à la conduite de nouveaux projets.

3.2.1. La dispersion du pendant politique du champ d'action stratégique de l'expérimentation sociale

Dans son premier rapport, le Conseil scientifique reconnaissait que la contrepartie du rythme effréné de la première année du FEJ était d'avoir construit un « *édifice fragile* » difficile à maintenir debout⁷⁷². En effet, si le FEJ arrive à se maintenir jusqu'à aujourd'hui, cette organisation échappe peu à peu au *champ d'action stratégique* l'ayant fondé, lui-même dissous par la perte de ses relais politiques.

Dans un premier temps, le Fonds d'expérimentation pour la Jeunesse est rattaché au ministère de la jeunesse et des solidarités actives, créé le 22 mars 2010 afin de remplacer les deux Hauts Commissariats dissous suite au départ du Haut Commissaire. Le ministre en charge, nouvellement entré au gouvernement, est Marc-Philippe Daubresse. C'est une période de transition pendant laquelle la Mafej est principalement occupée par le suivi des projets en cours, ou le lancement de ceux préparés avec le Haut Commissariat, comme par exemple le Revenu Contractualisé d'Autonomie.

Seul un appel à projets (AP4), intitulé « Favoriser le parcours des jeunes vers l'autonomie par les pratiques artistiques et culturelles », dit « Appel à projets culture », est ouvert le 8 juillet 2010⁷⁷³. Ces thématiques n'étaient pas abordées par le livre vert pour la jeunesse, et le FEJ s'écarte ainsi du protocole d'émergence des thématiques d'appel à projets définis à son origine. La clôture de l'appel à projets se fera en novembre 2010 et, sur les six projets retenus, aucun n'est évalué par la méthode expérimentale par assignation aléatoire. On remarque que les membres issus d'institutions scientifiques sont désormais peu présents dans le jury de sélection (7 sur 40), mais surtout que, par rapport aux appels à projets précédents, aucun membre du champ d'action stratégique et du conseil scientifique n'est présent⁷⁷⁴. Cet appel à projets est sans commune mesure avec les appels à projets précédents en France métropolitaine, puisque le FEJ n'engage que 629 098 euros pour les projets (6 projets au total) et 236 257 pour leurs évaluations⁷⁷⁵.

⁷⁷² CONSEIL SCIENTIFIQUE DU FEJ, *Rapport du Conseil scientifique du Fonds d'expérimentations pour la jeunesse, pour la période mai2009-décembre 2010*, Ministère de l'éducation nationale de la jeunesse et de la vie associative, 2011, p. 31.

⁷⁷³ Cet appel à projet propose trois thématiques : Mettre en cohérence les activités culturelles scolaires, péri et extra scolaires et réduire les inégalités relatives à l'accès aux offres culturelles, favoriser les pratiques artistiques et encourager l'expression des jeunes et favoriser l'insertion professionnelle par l'activité artistique et culturelle FONDS D'EXPERIMENTATIONS POUR LA JEUNESSE, *APPEL A PROJETS n°4*, « Favoriser le parcours des jeunes vers l'autonomie par les pratiques artistiques et culturelles », 8 juillet 2010.

⁷⁷⁴ CONSEIL SCIENTIFIQUE DU FEJ, *Rapport du Conseil scientifique du Fonds d'expérimentations pour la jeunesse, pour la période mai2009-décembre 2010*, op. cit. p. 43.

⁷⁷⁵ FONDS D'EXPERIMENTATION POUR LA JEUNESSE, *Rapport d'activité 2009-2011, Annexes*, 2012, p. 26.

Une autre expérimentation atypique préparée pendant la période du Haut Commissariat est lancée, « l'appel à projets quatre territoires » (APQT), sur les départements de Meurthe-et-Moselle, du Val d'Oise et du Morbihan, ainsi que la métropole de Nantes qui ensuite s'est retirée de l'expérimentation⁷⁷⁶. Huit projets sont sélectionnés, 1 133 344 euros sont engagés par le FEJ pour les projets et 163 768 pour l'évaluation⁷⁷⁷. Aucune des évaluations n'utilise la méthode expérimentale par assignation aléatoire.

Cette période de huit mois sous ce ministère est donc avant tout une période de gestion des projets en cours et de consolidation de la structure après l'élan impressionnant de la première année du Fonds. On observe néanmoins peu à peu une distanciation avec la raison d'être de cette institution, qui se manifeste par l'abandon des standards de l'évaluation par assignation aléatoire, et la moindre place accordée dans la sélection des projets au Conseil scientifique, au sein duquel sont toujours présents les membres scientifiques de l'ancien champ d'action stratégique. Le Fonds, conçu pour être indépendant des *champs d'action stratégiques* de l'administration centrale et des ministères, est peu à peu réinscrit dans la logique de ses ministères de tutelle. Le *champ d'action stratégique* de l'expérimentation sociale devait son existence à son indépendance des autres champs, par la maîtrise, notamment, de ses propres ressources, et par ses représentants politiques au cœur des instances étatiques. En effet, dans la théorie générale des champs, chaque champ est encadré dans un réseau d'autres champs et entretient une relation soit d'interdépendance soit hiérarchique. En présence de M. Hirsch, le champ d'action stratégique de l'expérimentation sociale subissait l'influence hiérarchique des autres champs de l'administration centrale uniquement via la participation de ses membres lors de la sélection des projets issus des appels à projets ; et avait réussi à s'en émanciper par les procédures hors appel à projets. Avec la suppression des Hauts Commissariats, ce champ perd d'une part, ses relais politiques et d'autre part, est placé sous la hiérarchie d'un champ d'action stratégique ministériel.

Lors du remaniement ministériel du 3^{ème} gouvernement Fillon en novembre 2010, le ministère de la Jeunesse et des Solidarités actives est supprimé, et le FEJ est alors rattaché au ministère de l'Éducation nationale, de la jeunesse et de la vie associative. Son ministre référent devient Luc Chatel, et la secrétaire d'État en charge de la jeunesse et de la vie associative, Jeannette Bougrab, prend en charge le Fonds. Les membres du champ d'action stratégique avec lesquels nous nous sommes entretenus évoquent cette période comme une période de crise par rapport à l'identité du FEJ, et notamment par rapport à sa vocation de développer l'évaluation expérimentale par assignation aléatoire. Un observateur interne constate : « *c'était l'Alpha et l'Omega de la politique de l'expérimentation notamment des expérimentations aléatoires* », puisque le budget du FEJ affecté à l'évaluation est remis en cause par la secrétaire d'État, qui s'interroge sur sa justification. Cela fait alors l'objet d'un argumentaire soutenant ces dépenses dans le premier rapport du Conseil scientifique produit à cette période, notamment pour les grands projets nationaux que l'on présente comme un investissement, afin de répondre aux critiques du ministère de tutelle⁷⁷⁸.

Ce rapport du Conseil scientifique du FEJ permet de dresser le bilan de cette période.

⁷⁷⁶ Cet appel à projets fut élaboré sur un diagnostic approfondi de la situation des jeunes de 16 à 25 ans sur chacun des territoires de juillet 2009 à janvier 2010, puis sur cette base des acteurs locaux ont proposé des projets qui furent co-financés par le FEJ

⁷⁷⁷ FONDS D'EXPERIMENTATION POUR LA JEUNESSE, *Rapport d'activité 2009-2011, Annexes*, 2012, p. 25.

⁷⁷⁸ « *C'est une somme importante, mais elle reste raisonnable en regard du coût des projets. Par exemple, l'évaluation des internats d'excellence coûte environ 160 000 euros par an pendant trois ans et demi : qu'est-ce en regard du coût de fonctionnement annuel des internats d'excellence de l'académie de Créteil et des 12 internats créés dans la foulée ? La subvention massive des permis de conduire (...) coûte déjà 10 millions d'euros : n'est-on pas prêt à dépenser 150 000 euros par an pour en mesurer précisément les bénéfices, avant d'orienter plus systématiquement des ressources vers ce type d'aide ? De toute évidence, l'évaluation est un investissement (dont la MAFEJ et le Conseil scientifique ont pour mission de garantir la qualité) : combien de centaines de millions d'euros sont dépensés sur des politiques rapidement abandonnées ou longuement décriées, et dont une évaluation préalable bien conçue, coûtant peut-être quelques centaines de milliers d'euros, aurait permis d'anticiper l'inefficacité ou l'inadaptation* » CONSEIL SCIENTIFIQUE DU FEJ, *Rapport du Conseil scientifique du Fonds d'expérimentations pour la jeunesse, pour la période mai2009-décembre 2010*, op. cit. p. 28.

L'« insistance » du Conseil scientifique sur l'expérimentation par assignation aléatoire y est rappelée tout en reconnaissant ses limites, mais en soulignant qu'elle est soutenue au plan politique⁷⁷⁹. Cette méthode est encore présentée comme la plus « fiable », mais néanmoins, comme toute méthode quantitative, ne permettant pas d'expliquer finement les raisons du succès d'un programme⁷⁸⁰. Ce rapport témoigne des premières difficultés rencontrées par l'utilisation non maîtrisée de cette méthode dans le cadre de l'Appel à projets 1, traitées dans le chapitre 3.

L'inscription du FEJ dans l'influence du champ d'action stratégique du ministère de l'Éducation se manifeste d'abord par la redéfinition des compétences de celui-ci, afin qu'il puisse être mobilisé pour les dispositifs de l'Éducation nationale. L'article 21 de la loi de finances rectificative du 29 décembre 2010 élargit le public du Fonds en le destinant non plus uniquement au 16-25 ans, mais aux moins de 25 ans, ainsi qu'à la réussite scolaire⁷⁸¹. Le Conseil scientifique relève aussi cette tendance dans leur rapport de 2001 : « *Lorsque le FEJ est passé sous la tutelle du Ministère de l'Éducation nationale, de la jeunesse et de la vie associative, la quasi-totalité des expérimentations soutenues a porté sur le champ scolaire, à la seule exception de l'appel à projets « Outre-Mer », dont l'un des thèmes était toutefois la lutte contre l'illettrisme* »⁷⁸². Le conseil scientifique regrette cette orientation et suggère d'aborder d'autres domaines d'expérimentations. Si les expérimentations changent sur le fond, elles se modifient aussi sur la forme⁷⁸³. Le Conseil scientifique est peu à peu évincé du processus d'élaboration et de sélection des appels à projets, détournant de fait l'instance de gouvernance auparavant favorable aux titulaires du champ d'action stratégique, afin de servir d'autres intérêts de champs. Le FEJ subit désormais la hiérarchie administrative, puisqu'il est inséré dans la Djepva, elle-même au sein du ministère. Dans ce contexte un des membres du Cabinet du Haut commissariat commente : « *ils ont juste été moins soutenus et moins scrutés, ce n'était plus très central alors qu'avant c'était central* ». Un observateur interne au FEJ relève l'ambiguïté de ce fonds qui a ses propres moyens de fonctionner, mais demeure néanmoins soumis à une tutelle administrative qui l'amène à conclure au détournement de cette institution : « *Le FEJ n'a pas été pensé comme ça* ».

Dans son rapport pour l'année 2011, le Conseil scientifique critique fortement les modalités d'élaboration des appels à projets qui sont au « *fil de l'eau* », « *sans que soit préalablement pensées les conditions de leur adaptation à l'approche expérimentale* », sans se référer à la « *littérature du champ* », et sans permettre au Conseil scientifique de discuter des modalités d'évaluation⁷⁸⁴. Les projets, beaucoup moins orientés vers l'émergence d'initiatives de terrain, sont plutôt définis par les pouvoirs publics, en particulier au sein de l'Éducation nationale⁷⁸⁵. Ils se rapprochent du modèle de procédure de l'appel à projets « permis de conduire », mais en laissant plus de marge de manœuvre aux porteurs de projets, et s'inscrivent dans les initiatives du département Recherche-développement, innovation et expérimentation (DRIE) de la DGSCO (Direction Générale de l'enseignement SCOLAIRE). Il

⁷⁷⁹ CONSEIL SCIENTIFIQUE DU FEJ, *Rapport du Conseil scientifique du Fonds d'expérimentations pour la jeunesse, pour la période mai2009-décembre 2010*, op. cit. p. 9.

⁷⁸⁰ Ibid. p. 10.

⁷⁸¹ « *Ce fonds est doté de contributions de l'État et de toute personne morale de droit public ou privé qui s'associent pour définir, financer et piloter un ou plusieurs programmes expérimentaux visant à favoriser la réussite scolaire des élèves et améliorer l'insertion sociale et professionnelle des jeunes de moins de vingt-cinq ans.* » Article 21 LOI n° 2010-1658 du 29 décembre 2010 de finances rectificative pour 2010, (2010) 2010-1658.

⁷⁸² CONSEIL SCIENTIFIQUE DU FEJ, *Rapport du Conseil Scientifique du Fonds d'expérimentation pour la jeunesse pour l'année 2011*, Fonds d'expérimentation pour la jeunesse, 2012, p. 16.

⁷⁸³ Un interlocuteur de la Mafej explique « il y a eu un moindre investissement, assez peu souvent avec un protocole d'évaluation aléatoire donc ça fait un vrai changement et de méthode et de thématique, c'est ce que j'appelle une réorientation »

⁷⁸⁴ CONSEIL SCIENTIFIQUE DU FEJ, *Rapport du Conseil Scientifique du Fonds d'expérimentation pour la jeunesse pour l'année 2011*, op. cit. p. 17.

⁷⁸⁵ Ibid. p. 4.

est important de noter que le directeur à l'époque de la DGSCO est Jean-Michel Blanquer, ancien recteur de l'académie de Créteil, qui n'est autre que l'ancien porteur de projet de l'expérimentation mallette des parents - la première expérimentation en milieu scolaire évaluée par la méthode expérimentale par assignation aléatoire – et soutenait activement cette démarche ⁷⁸⁶.

Alors que le FEJ avait été conçu en partie pour faire entrer de nouveaux acteurs scientifiques dans l'évaluation des politiques publiques, on observe à cette période un retour à l'évaluation interne des dispositifs par les services « études et évaluation » des ministères, et notamment par la Direction de l'évaluation, de la prospective et de la performance (DEPP) du ministère de l'Éducation. C'est dans ce cadre que sont lancés les appels à projets dits « APSCO ». L'APSCO1 « cours le matin, sport l'après-midi », engagés en juin 2011, prévoyait une évaluation unique des 58 projets par évaluation aléatoire, mais elle n'a pu être conduite faute des délais, trop restreints en amont du lancement du projet, a dû avoir recours par défaut à une évaluation qualitative ⁷⁸⁷. Sur les quatre appels à projets scolaires, les trois premiers font l'objet d'une évaluation centralisée qualitative, seul le quatrième présente une évaluation sur six ayant recours à la méthode aléatoire. On remarque aussi une nette baisse des fonds engagés.

Graphique 6: Montant des fonds engagés pour les évaluations et pour les projets par le FEJ en million d'euros pendant ses différentes périodes⁷⁸⁸



Lecture : Ce graphique montre que sous la période de M. Hirsch plus de 70% des fonds du FEJ en financement de projets ont été engagés et 80% des fonds totaux engagés en évaluation. Les expérimentations hors appel à projets sont aussi majoritairement amorcées par l'initiative de Martin Hirsch et affectent une somme équivalente aux projets et à leur évaluation. Les fonds engagés par les autres ministres sur les projets et leur évaluation représentent respectivement moins de 30% et moins de 10% de la somme intégrale dédiée à chaque catégorie.

Dans le rapport du conseil scientifique pour l'année 2011, le ton critique est manifeste. Ce dernier dénonce la non-utilisation de la méthode expérimentale par assignation aléatoire, et plus particulièrement le manque d'estimation d'impact dans les démarches d'évaluation, celles de l'APEDOM2, mais aussi le manque d'apprentissage du FEJ vis-a-vis de ses anciens appels à projets, comme par exemple entre l'appel à projets 1 et l'APSCO2. Puis, revenant sur

⁷⁸⁶ Cette posture du DGSO a été confirmée par cet interlocuteur lors d'un entretien ainsi que par un des membres du Conseil scientifique du FEJ lors d'un autre entretien : « J-M. Blanquer était à la djesco et soutenait cette démarche expérimentale. Ça (les expérimentations suivantes) vient plus de lui que du ministre ».

⁷⁸⁷ CONSEIL SCIENTIFIQUE DU FEJ, *Rapport du Conseil Scientifique du Fonds d'expérimentation pour la jeunesse pour l'année 2011*, Fonds d'expérimentation pour la jeunesse, 2012, p. 5.

⁷⁸⁸ Nous n'avons pas pu intégrer à cette comparaison le coût des expérimentations hors appel à projets puisque nous disposons uniquement de sa dotation globale sur toute la période du FEJ et non pas des dotations par évaluations par projets FONDS D'EXPERIMENTATION POUR LA JEUNESSE, *Rapport d'activité 2009-2011, Annexes*, 2012.

cette dynamique amorcée dès le deuxième semestre 2010 et l'extension des expérimentations d'internats d'excellences, évaluée par l'inspection générale, « *Le conseil scientifique regrette très vivement de n'avoir été à aucun moment consulté sur ce dossier. Nous aurions sans doute jugé que les exigences en matière d'évaluation qui ont accompagné cette vague de financement étaient très en deçà des attentes méthodologiques du FEJ.* »⁷⁸⁹. En outre, l'absence de publicisation de ce rapport amène le Conseil scientifique à caractériser cette démarche de « *particulièrement contraire à l'esprit du FEJ* ». Il stipule : « *Un financement substantiel a donc été engagé en contournant très largement les règles de fonctionnement du FEJ et surtout, sans organiser convenablement les conditions de l'évaluation.* »⁷⁹⁰. Cet écart de plus en plus important entre l'esprit d'origine du FEJ et son incarnation provoque même l'intervention de Martin Hirsch auprès de la secrétaire d'État et du ministre, afin de rappeler l'importance de la dimension évaluative du FEJ et de son financement⁷⁹¹. Il y a en effet, encore de grandes expérimentations à sécuriser, puisque c'est la période de lancement du RCA (revenu contractualisé d'autonomie), préparé pendant la période de Martin Hirsch, dont le principal volet auprès des missions locales engage 24 millions d'euros du FEJ, et l'évaluation menée par assignation aléatoire pour un montant d'environ 700 000 euros.

Dans ce même rapport, le Conseil scientifique déplore la pénurie, depuis le début du FEJ, d'équipes candidates pour l'évaluation, et appelle à un nouveau travail de communication auprès des chercheurs : « *ce déficit de candidatures interroge sur l'intérêt que suscitent les démarches d'évaluation des expérimentations sociales, dans le monde universitaire en particulier, et plaide pour une communication sur les enjeux et les objectifs de celles-ci* »⁷⁹²

La Mafej prend de l'ampleur. Initialement composée de trois personnes, cette équipe a peu à peu été renforcée et compte aujourd'hui 15 agents⁷⁹³. Seulement trois personnes composent le pôle « évaluation et capitalisation », ce qui amène le Conseil scientifique à émettre des inquiétudes quant à sa capacité à mener tout le travail de synthèse et de capitalisation, à alerter sur son besoin de ressources et d'un soutien politique fort, « *sinon l'investissement dans les expérimentations aura été parfaitement inutile* »⁷⁹⁴

Au final, à partir de 2009 jusqu'au changement de majorité présidentielle en mai 2012, le FEJ a été promoteur effectif de la méthode aléatoire principalement lors de la présence du Haut Commissaire et de son Cabinet, pour ensuite s'éloigner de cette voie suite à la double perte du soutien du Cabinet, et de l'influence du Conseil scientifique. Pendant ces trois ans, 33 % des projets ont été évalués par un protocole d'évaluation par assignation aléatoire. Étant donné qu'un grand nombre de projets ont fait l'objet d'une évaluation groupée dans le cadre des grands appels à projets nationaux, seuls 15% des évaluations menées ont suivi cette méthode⁷⁹⁵. Néanmoins, les projets évalués par celle-ci sont les plus importants et demeurent les plus emblématiques du FEJ. De fait, presque un quart des jeunes (23%) bénéficiant des projets du FEJ ont été concernés par ce type

⁷⁸⁹ Phrase soulignée en italique dans le texte CONSEIL SCIENTIFIQUE DU FEJ, *Rapport du Conseil Scientifique du Fonds d'expérimentation pour la jeunesse pour l'année 2011*, Fonds d'expérimentation pour la jeunesse, 2012, p. 10.

⁷⁹⁰ CONSEIL SCIENTIFIQUE DU FEJ, *Rapport du Conseil Scientifique du Fonds d'expérimentation pour la jeunesse pour l'année 2011*, Fonds d'expérimentation pour la jeunesse, 2012, p. 10.

⁷⁹¹ Élément précisé par Martin Hirsch et par un observateur interne du FEJ.

⁷⁹² FONDS D'EXPERIMENTATION POUR LA JEUNESSE, *Rapport d'activité 2009-2011*, 2012, p. 31; CONSEIL SCIENTIFIQUE DU FEJ, *Rapport du Conseil Scientifique du Fonds d'expérimentation pour la jeunesse pour l'année 2011*, Fonds d'expérimentation pour la jeunesse, 2012, p. 16.

⁷⁹³ Cette équipe est composée, d'un pôle administratif et financier (5 ETP), un pôle suivi et animation de projets (3 ETP), un pôle évaluation et capitalisation (3 ETP) et un pôle secrétariat-gestion du site internet (3 ETP), FONDS D'EXPERIMENTATION POUR LA JEUNESSE, *Rapport d'activité 2009-2011*, 2012, p. 11.

⁷⁹⁴ CONSEIL SCIENTIFIQUE DU FEJ, *Rapport du Conseil Scientifique du Fonds d'expérimentation pour la jeunesse pour l'année 2011*, op. cit. p. 17.

⁷⁹⁵ Nous voulons dire par là que plusieurs projets étaient évalués par la même évaluation aléatoire. Fonds d'expérimentation pour la jeunesse, *Rapport d'activité 2012* (Ministère des sports, de la jeunesse, de l'éducation populaire et de la vie associative, juillet 2013), 13.

d'évaluation, et environ la moitié (49%) des fonds engagés par le FEJ affectés à ces projets et à leur évaluation⁷⁹⁶. On observe donc à partir de cette période, une dissolution de l'alliance constitutive du champ d'action stratégique entre un groupe d'acteurs de l'action publique et un groupe d'acteurs scientifiques. Cette entreprise n'ayant pas mobilisé un nombre suffisant de suiveurs au sein des champs de l'action publique d'Etat et du champ scientifique, elle se retrouve alors limitée à des initiatives d'une part du groupe de titulaires composés des économistes statisticiens appliqués et d'anciens membres du Cabinet, maintenant dispersés dans l'administration centrale.

3.2.2. La subsistance du travail institutionnel cognitif suite à la dissolution du champ d'action stratégique

Avec la dissolution du champ d'action stratégique de l'expérimentation sociale, les anciens titulaires prennent à leur compte de continuer le travail institutionnel cognitif nécessaire à la bonne conduite de leurs projets en cours. Néanmoins, ces activités deviennent éparées et prennent la forme d'initiatives personnelles, ou bien d'alliances temporaires. Par ailleurs, elles doivent faire face à des initiatives de déstabilisations de plus en plus nombreuses.

Les chercheurs du Centre d'étude pour l'emploi, comme Yannick L'Horty et Pascale Petit, font un travail de légitimation de cette méthode dans deux articles, écrits en commun, *Evaluation aléatoire et expérimentations sociales* et *L'évaluation aléatoire : un succès qui ne doit rien au hasard*. Dans le premier, ils reviennent de manière détaillée sur les avantages de cette méthode, en se référant notamment aux travaux du prix Nobel J. Heckman, et la présente comme la méthode d'évaluation la plus fiable⁷⁹⁷. Cette rigueur scientifique, ainsi que l'investissement du chercheur que cette méthode implique, est plus loin convertie en vertu politique « *Elle est alors très lisible pour le policy maker qui va souvent lui accorder davantage de crédibilité que d'autres approches qui ne partagent pas ces propriétés.* »⁷⁹⁸. Ses limites sont évoquées, mais une réponse est apportée à chacune d'elles, particulièrement sur l'acceptabilité du tirage au sort, les effets d'équilibre et les biais de terrain, mais la méthode n'est pas remise en cause puisque les auteurs concluent que « *du point de vue de la productivité heuristique, l'évaluation aléatoire a largement fait ses preuves* »⁷⁹⁹. Cet article est une entreprise manifeste de légitimation scientifique et politique de cette méthode.

Un troisième article est publié par les mêmes auteurs sur laviedesidees.fr en mai 2011. Il a pour vocation d'expliquer le « succès » de cette méthode en France « *appliquées à des centaines de programmes sociaux locaux* »⁸⁰⁰. Ce succès semble ici surestimé puisque, certes, selon les chiffres du FEJ, 183 projets ont prévu d'être évalués par cette méthode, mais, d'une part, beaucoup de projets étaient regroupés autour de la même évaluation, ce qui rapporte le nombre d'évaluations aléatoires organisées par le FEJ à 42, et, d'autre part, plusieurs n'ont pas pu mener ce protocole à terme⁸⁰¹. Ce succès est expliqué, selon ces auteurs, par « *les qualités intrinsèques des méthodes d'évaluation aléatoire* » qui

⁷⁹⁶ FONDS D'EXPERIMENTATION POUR LA JEUNESSE, *Rapport d'activité 2009-2011*, 2012, p. 29.

⁷⁹⁷ Yannick L'HORTY et Pascale PETIT, *Evaluation aléatoire et expérimentations sociales*, 135, coll. Document de travail, Centre d'études de l'emploi, 2010, p. 15.

⁷⁹⁸ *Ibid.* p. 18.

⁷⁹⁹ *Ibid.* p. 23.

⁸⁰⁰ Yannick L'HORTY et Pascale PETIT, « L'évaluation aléatoire: un succès qui ne doit rien au hasard », *la vie des idées* (2011), p. 1.

⁸⁰¹ Nous avons calculé ces chiffres par rapport aux pourcentages d'expérimentations aléatoires indiqués dans FONDS D'EXPERIMENTATION POUR LA JEUNESSE, *Rapport d'activité 2009-2011, Annexes*, 2012.

sont en mesure de « *satisfaire pleinement les attentes des évaluateurs* »⁸⁰². Ces méthodes sont aussi présentées comme un vecteur de partenariat entre évaluateur et porteur de projet, pour une meilleure pertinence du travail de chacun. Néanmoins, cette démarche est propre à la plupart des méthodes d'évaluation employées pour évaluer des programmes innovants⁸⁰³. En conclusion le discours est modéré par la présentation de cette méthode comme étant une des ressources de la boîte à outils des chercheurs, mais non pas le « *Gold standard* » méthodologique, laissant la possibilité aux programmes non évaluables par cette méthode d'exister. Mais, pour un programme social local dont le public cible est de quelques centaines de bénéficiaires, elle est présentée dans bien des cas comme « *la seule approche quantitative adaptée* »⁸⁰⁴.

Un autre article de P. Zamora, datant de 2011 développe un argumentaire similaire en revenant sur l'intérêt scientifique et politique de la méthode qui ouvre « *un espace de communication et de coopération potentiellement extrêmement fécond entre scientifiques, praticiens et décideurs.* »⁸⁰⁵. Revenant sur l'impulsion de M. Hirsch il note « *l'intérêt politique* » pour cette méthode qui constitue une « *méthode de gouvernance qui permet de mieux objectiver les décisions et donc de mieux les faire accepter* »⁸⁰⁶. Le chercheur conclut cet article sur une note de modération, par une liste de précautions à prendre avant de décider de conduire cette méthode, et invite à rester attentifs aux effets externes d'un programme pouvant menacer la validité externe des résultats d'une expérimentation.

Par ailleurs, les chercheurs bénéficient de nouveaux financements d'Etat. L'École d'Économie de Paris est lauréate d'un financement par les « Grand Emprunt-Laboratoires d'excellence », et fonde le programme labex OSE (Ouvrir la Science Economique). Ils créent l'Institut des Politiques Publiques en 2012, lequel a pour objectif de promouvoir l'analyse et l'évaluation des politiques publiques. Il est financé par le Labex OSE et par les contrats d'études ou d'évaluations. Néanmoins, il défend son indépendance vis-à-vis de ces financeurs, notamment en publiant systématiquement les résultats produits. Cet institut privilégie les méthodes de micro-simulation, d'analyse coût-bénéfice, et « *les études d'impact s'appuyant sur des expériences naturelles ou randomisées* », en collaboration avec le J-PAL⁸⁰⁷. Ils continuent aussi à chercher des fonds de leur côté ; un économiste appliqué résume lors d'un entretien leur stratégie, après la dissolution du *champ d'action stratégique* : « *l'enjeu c'est de continuer, de faire, d'accumuler des preuves de l'intérêt de cette technologie pour des domaines d'application dans lesquels effectivement elle produit des effets tout à fait intéressants et donc de continuer à développer une offre.(...) Pour nous c'est un programme de recherche dans le cadre de notre fédération de recherche, et donc c'est un programme qui intéresse disons quelques dizaines de chercheurs dans quelques labos en France et qui sont de plus en plus intéressés par ces questions là. Nous, évidemment, on maintient, on développe, ça fait partie en quelque sorte de notre projet scientifique, académique, de développer ce type de méthodes, et si demain le Fonds d'Expérimentation pour la Jeunesse n'était pas là, on monte des projets comme ceux qu'on monte aujourd'hui en sollicitant d'autres... l'agence nationale pour la recherche, les fonds européens, les fonds publics, tous les financements de la recherche et les financements publics qui sont*

⁸⁰² Yannick L'HORTY et Pascale PETIT, « L'évaluation aléatoire: un succès qui ne doit rien au hasard », *op. cit.* p. 3-4.

⁸⁰³ Ce dialogue de co-construction entre porteur de projet et évaluateur fait l'objet du dernier de M. Q. Patton sur l'évaluation de projets expérimentaux innovants Michael Quinn PATTON, *Developmental evaluation : applying complexity concepts to enhance innovation and use*, New York, Guilford Press, 2010.

⁸⁰⁴ Yannick L'HORTY et Pascale PETIT, « L'évaluation aléatoire: un succès qui ne doit rien au hasard », *op. cit.* p. 7.

⁸⁰⁵ Philippe ZAMORA, « La méthode d'évaluation aléatoire: apports et limites », *Tracés* 2011.175- 186, p. 178.

⁸⁰⁶ *Ibid.* p. 180.

⁸⁰⁷ PARIS SCHOOL OF ECONOMICS, *La lettre: spéciale 5 ans*, n°8, janvier 2012.

disponibles pour mener des études, des recherches sur des thématiques qui intéressent le...qui intéressent. »

Du côté politique, des initiatives éparses subsistent grâce à la dispersion des anciens membres politiques du champ d'action stratégique dans les instances de l'administration centrale. Le Centre d'analyse stratégique (devenu Commissariat général à la stratégie et à la prospective le 22 avril 2013), avec La Direction Générale de la Cohésion Sociale, organise un cycle de séminaires « La cohésion sociale par ses instruments », dont la séance du 25 novembre 2010 est intitulée « l'expérimentation sociale innover par la preuve ? » est coorganisée avec l'Agence nouvelle des solidarités actives (Ansa). Une note préparée pour cet événement témoigne de la subsistance du travail de définition des membres du champ d'action stratégique. La définition de l'expérimentation sociale de l'Ansa est reprise et l'on entend par « méthode d'expérimentation sociale » la méthode scientifique aléatoire mobilisée pour évaluer des programmes innovants⁸⁰⁸. On observe ici la continuation du travail du champ d'action stratégique par ses anciens membres dans leur nouvelle position. Un des organisateurs et intervenants de cette séance n'est autre que Christophe Fourel, Chef de la Mission analyse stratégique, synthèses et prospective (MASSP) à la Direction Générale de la Cohésion Sociale (DGCS) qui, de 2007 à 2009, était Directeur général de l'Ansa, et qui promeut la méthode aléatoire⁸⁰⁹. D'autres membres de l'ancien champ d'action stratégique interviennent : un membre politique, Étienne Grass, ancien directeur de cabinet de Martin Hirsch, qui est depuis conseiller social à la représentation permanente de la France auprès de l'Union européenne et soutient que « *la méthode d'évaluation la plus rigoureuse repose sur la randomisation* »⁸¹⁰ ; un membre scientifique, Yannick l'Horty, qui « *considère que cette méthode est la plus rigoureuse des méthodes d'évaluations quantitatives* » et précise que « *l'évaluation doit rester plurielle évidemment. Il faut des techniques qualitatives et des techniques quantitatives parce que si on ne fait que mesurer des chiffres parfois on aura du mal à interpréter les causalités ou interpréter ce qu'on essaie de mesurer. Mais néanmoins, il y a aussi de la hiérarchie à l'intérieur des méthodes quantitatives.* » Le travail de promotion de l'expérimentation aléatoire est donc entretenu par les anciens membres du Cabinet de Martin Hirsch. Par exemple, Étienne Grass porte la promotion au niveau européen, comme évoqué plus haut, et Jérôme Teillard renforce l'ancrage des expérimentations au Ministère de l'éducation amorcé par J.M Blanquer et Bénédicte Robert en intégrant le département recherche-développement, innovation et expérimentation à la DGESCO.

Esther Duflo gagne en notoriété en recevant la médaille Clark en avril 2010. Le prestige acquis auprès du gouvernement français, et notamment du Ministère de l'enseignement supérieur et de la recherche amène le ministre à lui remettre la médaille de l'innovation le 28 avril 2011⁸¹¹. Elle publie, en 2010, l'ouvrage *Le développement humain. Lutter contre la pauvreté*, et au cours de la promotion de cet ouvrage elle continue à soutenir la dynamique du FEJ « *ce fonds constitue un exemple pour le monde. La France est, pour une fois, très*

⁸⁰⁸ DIRECTION GENERALE DE LA COHESION SOCIALE, ANSA, *L'expérimentation sociale: innover par la preuve?*, Centre d'Analyse Stratégique, 25 novembre 2010.

⁸⁰⁹ En témoigne son commentaire favorable de l'ouvrage d'Esther Duflo et Abhijit Banerjee, *Repenser la pauvreté*, http://www.nonfiction.fr/article-5574-des_methodes_scientifiques_pour_lutter_contre_la_pauvrete.htm

⁸¹⁰ AGENCE NOUVELLE DES SOLIDARITES ACTIVES, *Séance 2- L'expérimentation sociale: innover par la preuve? Programme*, 25 novembre 2010.

⁸¹¹ <http://www.enseignementsup-recherche.gouv.fr/cid55906/medaille-de-l-innovation-a-esther-duflo-mathias-fink-et-francois-pierrot.html>

nettement en avance par rapport à ses homologues européens. »⁸¹² Un ouvrage écrit avec Abhijit Banerjee, *Repenser la pauvreté*, paraît en 2012, et contribue à légitimer la pertinence de la méthode aléatoire.

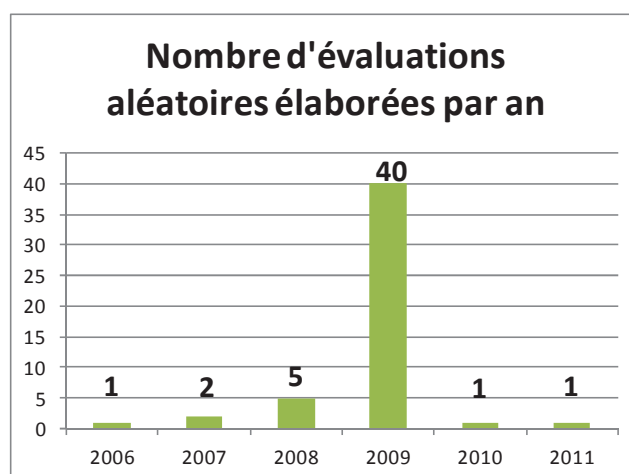
De son côté Martin Hirsch relaie dans son blog les résultats d'expérimentations aléatoires du FEJ et intervient, si nécessaire, comme nous l'avons vu précédemment, lorsque le Conseil scientifique du FEJ est en difficulté.

À partir de 2010, les critiques de l'expérimentation par assignation aléatoire deviennent plus nombreuses et plus organisées. Elle émerge d'économistes hétérodoxes, qui connaissent l'impact de cette méthode dans le champ stratégique de l'économie du développement. On note l'article d'Agnès Labrousse, de 2010 qui revient entre autres sur le processus de transfert technologique de cette méthode du champ de la médecine aux politiques de développement. Elle attire l'attention sur le fait que, comme toute autre pratique scientifique, cette méthode relève d'une construction sociale, de bricolages sur le terrain, dont il faut être conscient, et qu'elle ne peut constituer un substitut à d'autres méthodes d'analyse économique⁸¹³.

Le groupe « standard et déontologie » de la Société Française d'Evaluation propose aussi de décroiser le débat sur l'évaluation d'impact en soulignant notamment la rigueur et la pertinence de l'emploi d'autres méthodes d'évaluations pour estimer l'impact des projets expérimentaux, et refuse une hiérarchie des méthodes. Il organise une journée d'étude, en octobre 2010, lors de laquelle il fait intervenir Marc Gurgand de l'École d'Économie de Paris, mais aussi d'autres intervenants proposant des méthodes alternatives. Cette journée donne lieu à la rédaction d'un cahier resituant l'expérimentation aléatoire parmi un ensemble d'autres méthodes d'évaluation, et en souligne les limites⁸¹⁴.

Au final, en 2012, l'intégralité des acteurs avec lesquels nous nous sommes entretenus s'accordent à évoquer la période du Haut Commissariat comme l'âge d'or de l'expérimentation aléatoire en France et reconnaissent que « *le souffle est retombé* ».

Graphique 7: Nombre d'évaluation aléatoires élaborées par an sur le territoire français



Lecture : Il s'écoule beaucoup de temps entre l'élaboration d'une expérimentation et son lancement, nous avons choisi de classer les expérimentations par année selon le moment soit de leur sélection par un appel à projets soit si elles ne sont pas issues d'un appel à projets de leur élaboration. Il y a une

⁸¹² Esther DUFLO et Claire MONTIALOUX, « 4. Repenser la lutte contre la pauvreté », *Regards croisés sur l'économie* 2011.10.59- 62.

⁸¹³ Agnès LABROUSSE, « Nouvelle économie du développement et essais cliniques randomisés: une mise en perspective d'un outil de preuve et de gouvernement », *op. cit.*

⁸¹⁴ SOCIÉTÉ FRANÇAISE D'EVALUATION, *L'évaluation des impacts des programmes et services publics*, Cahier de la SFE.

grande porosité entre les années 2009 et 2010 pour lesquels de nombreux projets ont été élaborés à cheval sur ces deux années. Nous précisons que nous comptons bien en nombre d'évaluations et non pas en nombre de projets évalués. Ces chiffres sont aussi évalués à la hausse puisque plusieurs évaluations ne sont pas parvenues à terme et ont dû soit être interrompues soit mobiliser un autre protocole. Ils montrent néanmoins une nette concentration de l'élaboration et du financement d'expérimentations aléatoires en France par le FEJ lors de la présence de M. Hirsch. Sans cet apport de financements la diffusion de cette méthode aurait été beaucoup plus marginale.

L'institutionnalisation de cette méthode d'évaluation comme la plus rigoureuse au sein des instances d'expertise d'État a été fulgurante et résulte d'un travail intensif des membres du champ d'action stratégique afin de diffuser la « culture expérimentale ». Néanmoins, le pendant politique de ce champ d'action stratégique étant restreint aux Cabinets de M. Hirsch, il n'a pu se maintenir sans un entrepreneur à même de faire le lien avec les instances politiques supérieures. Nous reviendrons dans l'épilogue de ce travail sur un état des lieux de la diffusion de cette méthode avec le changement de majorité présidentielle en 2012.

Conclusion du chapitre 2

Le questionnement de cette partie était double. Il consistait, d'une part, à identifier et à expliquer les conditions de l'établissement aussi rapide de la légitimité de la méthode d'évaluation expérimentale par assignation aléatoire et, d'autre part, à rendre compte de l'impact de ce développement sur les modalités d'interaction du champ de l'économie appliquée et de l'action publique en France.

Nous avons dans un premier temps, identifiés les motivations du développement de la méthode expérimentale en France. Cette méthode a permis de répondre à plusieurs crises. Notamment celles relevant du champ de l'économie appliquée. Cette crise était double, elle consistait à un manque de financement, et de notoriété des économistes appliqués français sur la scène internationale. La méthode par assignation aléatoire une fois reconnue par l'administration centrale, a ouvert l'accès à de nouveaux financements tout en permettant la production de résultats d'évaluation aléatoire qu'ils ont pu mobiliser pour participer aux débats scientifiques internationaux se fondant sur des résultats issus d'expérimentations.

Dans le champ de l'action publique la crise est aussi multiple. En temps de crise économique, l'État a pour préoccupation constante de maîtriser ses dépenses. Dans ce contexte, son intervention rencontre une crise de légitimité au regard de son efficacité. Dans un contexte de diffusion des politiques d'activation des aides sociales ou des politiques de l'emploi, la question de l'efficacité des aides sociales à inciter le changement de comportement des individus est très prégnante. Cette méthode permet d'afficher une neutralité idéologique par la mobilisation de l'objectivité scientifique, et de s'inscrire dans une rhétorique d'intervention jugée au regard de son efficacité. Elle permet aussi d'amorcer de manière « temporaire » des réformes non populaires par le biais d'expérimentations, auprès d'une opinion publique sceptique. Néanmoins cette rhétorique a été surtout mobilisée par le Cabinet de M. Hirsch, afin de justifier l'action d'un acteur sympathisant de gauche au sein d'un gouvernement de droite, en menant des politiques sociales, non pas guidé par l'idéologie, mais par un souci d'efficacité. Le champ d'action stratégique de l'expérimentation sociale a donc été construit par une partie des membres de ces champs en crise, lesquels, par ce moyen, pouvaient se saisir de nouvelles ressources et gagner en légitimité.

Ces deux impulsions issues de deux champs distincts se retrouvent dans l'élaboration d'un même discours sur l'expérimentation sociale regroupant la conduite de projets

innovants et leur évaluation par la méthode expérimentale par assignation aléatoire. Ce discours procède par un discrédit des modes d'expertise précédents, ainsi que des autres méthodes d'évaluation. On voit que par un travail institutionnel cognitif, de réseaux, et d'accaparement de ressources, le champ d'action stratégique émergent procède à une tentative de son institutionnalisation au sein des instances Etatiques.

Cette méthode, contrairement aux autres méthodes d'évaluations, semble de prime abord avoir rencontré un succès rapide auprès de l'administration centrale. En effet, la méthode aléatoire n'est pas un instrument qui menace les politiques existantes, puisque son emploi est restreint aux politiques innovantes, elle s'inscrit dans la lignée des outils de prospective développés par les experts économiques de l'Etat Français. Néanmoins, cette méthode modifie la nature de certains rapports traditionnels entre ces deux champs. La nature scientifique de cette méthode ouvre la possibilité d'une porosité de financements entre la subvention de la recherche et la commande d'évaluations par l'Etat. Elle amène de fait une ambiguïté entre l'apport du résultat de l'expérimentation pour la recherche et pour la décision publique – données fiables pour le scientifique, légitimité ou utilité pour le politique –, nous étudierons cet équilibre au chapitre 3. Cette méthode positionne le chercheur comme « co-expérimentateur », il lui permet même parfois d'être à l'initiative de projets expérimentaux, ce qui lui confère une place de choix dans la programmation des modalités de l'action publique et de la mise en agenda de nouveaux types d'intervention ainsi que la définition de leurs objectifs. La promotion de cette méthode a aussi amené en retour, les commandes d'Etat à intervenir auprès des scientifiques dans le choix de leur méthodologie, du fait qu'ils favorisaient le financement des projets évalués par cette méthode.

Cette méthode doit surtout son développement rapide grâce à sa promotion par les acteurs traditionnels de l'expertise d'Etat, c'est à dire les économistes appliqués issus de l'école de l'Ensaë. Ces statisticiens ont bénéficié de relais privilégiés au sein des Services Statistiques Ministériels qui disposaient d'une culture scientifique similaire de par leur formation commune. Néanmoins, cette proximité avec l'action publique étant restreinte à la Dares, la conduite d'un nombre d'expérimentations aléatoires plus conséquent a dû passer par une alliance avec d'autres acteurs de l'action publique. Ce travail a été mené par une poignée d'entrepreneurs appartenant soit au champ de l'économie appliquée soit à celui de la haute administration et un entrepreneur frontière, Martin Hirsch. Les qualités d'entrepreneur frontière de ce dernier, ont été indispensables à l'élaboration d'un champ d'action stratégique mobilisant des acteurs de l'action sociale, des chercheurs, des acteurs politiques et des fonctionnaires. Par ses multiples facettes, et réseaux, il a été le catalyseur du mouvement d'institutionnalisation de cette méthode. Il a su faire des concessions aux moments opportuns, au risque parfois de briser cette alliance, et il a joué en permanence sur la double présentation de la méthode expérimentale, soit comme méthode scientifique, soit comme instrument à même de faire émerger des innovations sociales du terrain. Le couplage de ces deux démarches a été l'objet du travail cognitif intense organisé dans les champs adjacents à ce nouveau champ stratégique afin d'attirer le maximum d'acteurs et de ressources. Cependant, la tentative d'institutionnalisation de ce champ d'action stratégique, par son inscription dans l'administration centrale n'a connu qu'un succès temporaire. D'abord, la tentative d'élargissement de ce champ d'action stratégique par la procédure d'appel à projets s'est avéré être un échec. Cette stratégie d'intéressement s'est vue détournée par les membres de champ stratégiques interreliés à ce dernier qui sont intervenus dans les unités de gouvernance interne, ici incarnés par les jurys des appels à projets. L'ensemble de titulaires scientifiques et politiques n'ont pas réussi à maîtriser la raison d'être de ce champ, à savoir, la promotion de la méthode expérimentale par assignation aléatoire, de

par le manque de « suiveurs » autant dans les différents champs de l'action publique que dans ceux des sciences sociales. Ils ont rencontré d'une part des *challengeurs compétents* issus des champs stratégiques de l'action publique qui ont imposé leurs critères traditionnels d'allocation de ressources financières allant à l'encontre du nouveau modèle défendu par le champ d'action stratégique de la méthode expérimentale. Ils ont du faire face aussi à un manque de *suiveurs* dans le champ des sciences sociales appliquées à même de maîtriser cette nouvelle méthode d'évaluation. On observe donc l'émergence d'un champ d'action stratégique peinant à élargir ses membres au-delà des *titulaires* ayant participé à sa promotion et ce malgré l'ampleur de leur travail institutionnel. Le départ de cet entrepreneur frontière et de son équipe de gouvernement marque alors la dissolution de ce champ hybride par la dispersion de ses relais politique au sein de l'administration centrale. Cette structure affaiblie est devenu alors la cible de *challengeurs* compétents réinstallant la logique de champs adjacents en leur faveur et aux dépens des anciens titulaires.

Néanmoins, si cette alliance semble se dissoudre à partir de 2010, une autre interaction entre acteurs de l'action publique et de l'économie appliquée s'est organisée par ailleurs. Cette interaction est d'ordre micro-sociologique et se structure selon différentes modalités entre les parties prenantes de chaque expérimentation sociale évaluée par la méthode expérimentale par assignation aléatoire. C'est cette pluralité d'interactions que nous allons étudier dans le chapitre suivant.

***3^{ème} Partie : La méthode expérimentale par
assignation aléatoire et ses pratiques :
étude des tensions d'apprentissages
d'une relation clinique***

« Chaque expérimentation est une histoire singulière (...). Il faut arbitrer entre (...) un idéal pur pour l'évaluateur et des contraintes qui sont posées et qui peuvent parfois limiter la pertinence du programme du point de vue d'un expérimentateur ; et inversement un programme très pertinent n'est pas forcément un programme facilement évaluable. L'intérêt, bien sûr, c'est cette rencontre entre les contraintes, les souhaits de l'expérimentateur et ceux de l'évaluateur, c'est presque comme une histoire d'amour, comme toutes les rencontres on peut avoir le pire et le meilleur, c'est pour ça que je pense que c'est singulier. On a fait la sociologie de la famille et du mariage, et on pourrait faire la sociologie des rencontres expérimentateur-évaluateur avec parfois des divorces, des choses qui se passent mal à cause de l'évaluateur, d'autres qui se passent mal à cause de l'expérimentateur. ».

Propos issus d'un entretien avec un économiste statisticien membre du Conseil scientifique du Fonds d'Expérimentation pour la Jeunesse

« Celui qui n'a pas le moyen d'influencer d'une façon ou d'une autre la représentation des faits, n'existe pas dans le monde politique »

Jacques ELLUL, *L'illusion politique*, Paris, Table ronde, 1965, p. 172.

L'étude de la promotion de la méthode expérimentale par assignation aléatoire en France, effectuée dans le chapitre précédent, appelle une investigation sur les modalités de l'incarnation de cette pratique sur le territoire français. Ce dernier chapitre déplace la focale d'attention à un niveau microsociologique afin d'étudier l'interaction de la science et de l'action publique organisée par l'expérimentation sociale évaluée par cette méthode. Nous souhaitons ainsi confronter l'argumentaire de promotion de la méthode aléatoire à sa pratique effective et ainsi nourrir le débat environnant cette méthode, grâce à une étude approfondie de ses incarnations sur le terrain.

Ce chapitre s'inscrit dans notre questionnement transversal sur l'impact de cette méthode sur la reconfiguration des interactions entre les sciences sociales et l'action publique en France. Nous nous intéresserons ici particulièrement aux variables témoignant de cette relation et l'influençant. Il est nécessaire de souligner que ce choix a émergé directement de notre travail de terrain. L'ensemble des acteurs rencontrés ont en effet évoqué la dimension structurante de cette interaction dans l'expérimentation aléatoire.

Ce chapitre introduit un second questionnement propre à l'étude de cet objet grâce à cette échelle microsociologique. Cette observation de *l'evidence-based policy* en pratique nous offre une fenêtre d'accès aux mécanismes d'apprentissages des acteurs scientifiques et ceux des acteurs publics, organisés par la démarche expérimentale aléatoire. L'ambition affirmée de cette méthode d'évaluation est d'organiser la rencontre des intérêts et processus d'apprentissage de ces deux parties. Nous nous demanderons donc dans quelle mesure la méthode expérimentale par assignation aléatoire, présentée comme « l'étalon or » de l'apprentissage scientifique, parvient à organiser l'apprentissage des acteurs de l'action publique. En d'autres termes nous étudierons l'utilité pratique de cette dernière pour ce groupe d'acteurs.

Nous sommes déjà revenus dans le Chapitre 1 sur le cadrage théorique de ce chapitre. Nous mobilisons ici à nouveau le concept de *champ d'action stratégique* ainsi que celui de *travail institutionnel*, en les ayant enrichis par les réflexions propres aux instruments d'apprentissage. Ce travail théorique a conduit à l'émission de quatre hypothèses répondant aux deux questionnements évoqués ci-dessus. Ces hypothèses sont les suivantes :

- Hypothèse 1 : La pratique de cette méthode organise de nouveaux types d'interactions entre les acteurs de l'action publique et les milieux scientifiques, mais ces interactions sont variables d'un site institutionnel à l'autre selon l'équilibre atteint entre les différentes activités de *travail institutionnel* en leur sein.
- Hypothèse 2 : Les enjeux liés au *powering* (compétition pour l'accès aux ressources) au sein de l'expérimentation aléatoire conditionnent les activités de *puzzling* (résolution des problèmes publics) pour chaque partie investie dans une expérimentation aléatoire et s'inscrivent dans les logiques de champs d'action stratégiques d'appartenance des parties prenantes.
- Hypothèse 3 : Le processus de production de connaissance de la méthode aléatoire constitue un obstacle à l'apprentissage des acteurs de l'action publique pendant la conduite de l'expérimentation

- Hypothèse 4 : L'apprentissage recherché par les chercheurs dans les résultats produits par la méthode aléatoire ne correspond pas aux types de connaissances nécessaires à l'apprentissage des acteurs de l'action publique.

Afin d'étudier ces hypothèses cette partie a recours à la conduite et l'analyse d'études de cas. Avant de présenter nos résultats empiriques, nous proposons de présenter dans une partie introductive les cas d'expérimentation sélectionnés.

Les trois parties suivantes de ce chapitre sont consacrées à la restitution de l'analyse des données de terrain au prisme de cet appareillage méthodologique. Nous avons découpé cette étude de l'expérimentation en trois temps dont chacun constitue une partie.

D'abord nous abordons la phase de conception de l'expérimentation. Nous revenons sur les raisons ayant motivé les acteurs scientifiques et ceux de l'action publique à s'engager dans la démarche expérimentale, en les restituant dans les enjeux de leurs champs stratégiques respectifs. L'étude de ces intérêts d'apprentissage, parfois communs, parfois divergents, confirme la conceptualisation de l'expérimentation aléatoire comme une *relation clinique* dans laquelle chaque partie prenante poursuit sa propre entreprise. L'étude de la rencontre de ces parties prenantes ainsi que le design des expérimentations donne à voir des types de *travail institutionnel* contribuant à forger différentes incarnations de l'institution de l'expérimentation aléatoire, des différences qui s'expliquent notamment par les compromis effectués par ces acteurs.

Cet équilibre entre les intérêts de ces parties prenantes s'incarne ensuite dans la phase de mise en œuvre des expérimentations. Celle-ci témoigne du *travail institutionnel* le plus conséquent. Pour les chercheurs, le contrôle du traitement ainsi que celui du tirage aléatoire sont nécessaires afin de garantir l'apprentissage scientifique. Cependant, ces exigences de contrôle entrent souvent en opposition avec les pratiques des acteurs de la mise en œuvre, lesquels peuvent alors aller jusqu'à s'investir dans un travail de déstabilisation institutionnelle de cette méthode et compromettre ainsi la production de résultats.

Enfin, nous étudions la phase de production et d'analyse des résultats ainsi que leur utilisation par les chercheurs et les porteurs de projet et les financeurs. Les difficultés rencontrées lors de la phase de mise en œuvre contribuent alors à endommager la production d'estimations rigoureuses de l'impact de la politique publique par cette méthode, et les chercheurs ont alors souvent recours à d'autres méthodes d'analyse ou de collecte de données afin de consolider leurs résultats. Néanmoins, par un travail poussé de capitalisation scientifique, la plupart des chercheurs atteignent finalement leurs objectifs d'apprentissage et parviennent à valoriser cette expérience dans leurs travaux académiques. Du côté des acteurs de l'action publique, plusieurs éléments font obstacle à l'utilisation des résultats produits, notamment une disponibilité des résultats trop tardive ainsi qu'un éloignement de l'objet de ces résultats des intérêts d'apprentissage de ces acteurs. Néanmoins, quelques cas témoignent d'un maintien d'une *relation clinique* autour d'un intérêt commun des parties prenantes par une convergence particulière de mise à l'agenda politique et de recherche académique.

Pour conclure, nous reviendrons sur les particularités de cette méthode d'évaluation et de son emploi dans le cadre de l'expérimentation sociale, tout en redéfinissant, à la lumière de ce travail empirique, les conditions pertinentes de son utilisation.

1. Présentation des études de cas conduites et analysées dans ce chapitre

Nous avons déjà précisé dans le premier chapitre de ce travail notre stratégie de sélection des cas pour cette étude (voire chapitre 1 2.2.2.1). Cette stratégie nous a amené à sélectionner au final 15 études de cas. D'autres expérimentations pouvaient entrer dans ces critères mais nous les avons écartées faute de disponibilité des données ou parce que bien qu'affichant un recours à la méthode aléatoire, il apparaissait au final qu'elles s'apparentaient davantage à des méthodes quasi-expérimentales, ou encore parce que leur rapports finaux d'évaluation ont été rendus publics trop tardivement pour entrer dans notre analyse⁸¹⁵.

Le tableau ci-dessous présente les 15 cas sélectionnés en détaillant l'intitulé de chaque expérimentation, son commanditaire/financeur, le porteur de projet, le laboratoire effectuant l'évaluation, et la période d'expérimentation.

⁸¹⁵ Entrent dans cette catégorie : Nouveau dispositif d'orientation en première année de licence (Apdiises 12), AP1 43 pass accompagnement, Ap1 348 un pas dehors, Ap1 112 tulle 2 : Promotion de l'apprentissage et sécurisation des parcours des jeunes apprentis en France, AP1 440 Développement de partenariats avec des acteurs professionnels de l'insertion et du placement, HAP2 Coup de pouce clé, HAP3 créa jeune.

Table 3 : Tableau récapitulatif de l'ensemble des cas étudiés dans ce chapitre

Intitulé de l'expérimentation	Commanditaire de l'expérimentation	Porteur de projet	Evaluateur	Période de mise en œuvre de l'expérimentation
1. Accompagnement renforcé des bénéficiaires du RMI	Conseil Général des Hauts de Seine	Conseil Général des Hauts de Seine, prestataire privé	Crest-Dares	mars 2006 - oct 2007
2. Accompagnement des jeunes diplômés demandeurs d'emploi par des opérateurs privés de placement (jeunes diplômés)	DGEFP	DGEFP sous traité à prestataire privé	Crest-Dares-MIT	sept 2007 – juin 2009
3. Accompagnement des demandeurs d'emploi par les opérateurs privés de placement et le programme Cap vers l'entreprise (OPP-CVE)	Unédic/Anpe puis Pôle Emploi	Unédic/Anpe puis Pôle Emploi	Crest-Ecole d'économie de Paris (PSE)	dec 2007 – juin 2008
4. Programme de prévention des ruptures dans l'apprentissage (Tulle) (AP Diées 09)	Haut Commissariat aux Solidarités Active 1 ^{er} appel à projets	Mission Locale de Tulle	Crest-Dares	avril 2008 - dec 2011
5. Mallette des parents 6 ^{ème} (AP Diées 11)	HCSA 1 ^{er} appel à projets	Rectorat de Créteil	PSE	juil 2008 - juil 2010
6. Orientation des jeunes au lycée via des dispositifs de parrainage (Actences) (AP Diées 01)	HCSA 1 ^{er} appel à projets	Association Actences	PSE	sept 2008 – juin 2011
7. Internat d'excellence de Sourdon (HAP 1)	Haut Commissariat à la Jeunesse hors appel à projets	Rectorat de Créteil	PSE	août 2009 – dec 2012
8. Tous en Stage (AP1 464)	HCJ 1 ^{er} appel à projets	Association Réussir moi aussi	Tepp, Université d'Evry	sept 2009 – juin 2011
9. Améliorer la Qualité des emplois exercés par les étudiants (AQ3E) (AP1 336)	HCJ 1 ^{er} appel à projets	Université du Maine	Tepp, Univ. Paris-est Marne la Vallée	sept 2009 – juin 2012
10. Inscrire les contrats en alternance dans une logique de parcours sécurisé (Beaune) (AP1 263)	HCJ 1 ^{er} appel à projets	Mission Locale de Beaune	Iredu Cereq	oct 2009-juil 2011
11. CV anonyme	Commissariat à la diversité et à l'égalité des chances	Sous traité à Pôle emploi	Crest-PSE-sous dir. des études de Pôle Emploi	nov 2009- nov 2010
12. Rebond pour les étudiants décrocheurs de l'enseignement supérieur (AP1 317)	HCJ 1 ^{er} appel à projets	Association AFIJ	Crest	dec 2009 – dec 2011
13. Accompagnement partenarial d'itinéraires personnalisés d'accès à la qualification (APIPAQ) (AP1 359)	HCJ 1 ^{er} appel à projets	Mission Locale de Marseille, et mission d'insertion de l'éducation nationale	Cereq	dec 2009- dec 2011
14. « 10 000 permis pour réussir »	HCJ	Plusieurs Missions Locales	Tepp	janv 2010- sept 2010
15. AgiLoJe	HCJ 1 ^{er} appel à projets	Missions Locales et Associations	Credoc	juil 2010 –nov 2011

Ce travail mobilise les données issues de ces cas au cours du développement de ce chapitre, néanmoins afin de faciliter le repérage du lecteur parmi la diversité de cas, nous souhaitons procéder à une présentation succincte de ces derniers. Est présenté ici un descriptif de l'expérimentation proposée, le cadre de son financement, le porteur de projet, l'équipe d'évaluation, et la période de mise en œuvre.

1.1.1.1. Accompagnement renforcé des bénéficiaires du RMI

Financier de l'évaluation : Conseil Général des Hauts de Seine

Porteur du projet : Conseil Général des Hauts de Seine, sous traité à l'opérateur privé Ingeus

Equipe d'évaluation : Bruno Crépon (Crest-Insee et Jpal), Marc Gurgand (Paris School of Economics, Crest-Insee et Jpal), Thierry Kamionka (Crest-Insee), Laurent Lequien (Dares et Crest-Insee)

Période de mise en œuvre de l'expérimentation : mars 2006 - oct 2007

Remise du rapport final d'évaluation: non précisé

Présentation de l'intervention : Accompagnement des RMistes les plus éloignés de l'emploi par des opérateurs privés. L'accompagnement consiste à organiser des réunions régulières, une par semaine, avec un conseiller unique qui aide les personnes à gagner en confiance, à résoudre des problèmes sociaux immédiats et à chercher du travail ou une formation. L'objectif du dispositif est le retour à l'emploi.

Résultats de l'évaluation : La puissance statistique est trop faible pour avoir des résultats précis. Est repéré une faible amélioration de l'accès à l'emploi qui réduit un peu la dépendance à l'aide sociale. Le programme est plus efficace pour ceux qui bénéficient depuis peu de temps de l'aide sociale. Cependant, le coût de l'intervention (paiement au fournisseur privé) est suffisamment élevé pour que le bénéfice net soit clairement négatif

Généralisation ou pérennisation de l'expérimentation : oui

1.1.1.2. Accompagnement des jeunes diplômés demandeurs d'emploi par des opérateurs privés de placement (jeunes diplômés)

Financier de l'évaluation : DGEFP-Dares

Porteur du projet : DGEFP sous traité à opérateur privé

Equipe d'évaluation : Bruno Crépon (Crest-Insee et Jpal), Esther Duflo (Jpal MIT), Marc Gurgand (Paris School of Economics, Crest-Insee et Jpal), Roland Rathelot (Dares Crest-Insee), Philippe Zamora (Dares et Crest-Insee)

Période de mise en œuvre : sept 2007 – juin 2009

Remise du rapport final d'évaluation : pas de rapport en tant que tel, publication dans *Dares analyse* en sept 2011

Présentation de l'intervention : Accompagnement renforcé vers l'emploi de jeunes de moins de 30 ans demandeurs d'emploi, de jeunes de moins de 30 ans au chômage depuis plus de 6 mois, niveau Bac plus 2 minimum, par des opérateurs privés de placement (OPP) des secteurs marchands et non marchands. Suivi d'un accompagnement des jeunes dans leur emploi pendant 6 mois. L'objectif du dispositif est le retour à l'emploi et le maintien dans l'emploi.

Résultats : À l'horizon de 8 mois, l'accompagnement par les OPP pris dans leur

ensemble a augmenté significativement la proportion de jeunes occupant un CDD d'au moins 6 mois. À un horizon plus lointain (16 ou 20 mois), l'accompagnement par les OPP n'a plus eu d'impact statistiquement significatif sur le taux d'emploi comme sur le taux d'emploi durable, ce qui laisse à penser que l'intervention des OPP a principalement eu pour effet d'accélérer l'accès à l'emploi des bénéficiaires. Par ailleurs, l'évaluation a mis en évidence des effets d'éviction à court terme : pour certains métiers recherchés essentiellement par des jeunes diplômés, l'accès plus rapide à l'emploi durable des jeunes pris en charge par les OPP s'est effectué en large partie au détriment des jeunes non accompagnés.

Généralisation ou pérennisation de l'expérimentation : non

1.1.1.3. Accompagnement des demandeurs d'emploi par les opérateurs privés de placement et le programme Cap vers l'entreprise (OPP-CVE)

Financier de l'évaluation : Unédic-ANPE puis Pôle emploi.

Porteur du projet : Unédic, sous traité à des opérateurs privés -ANPE puis Pôle emploi sous traité à des opérateurs privés

Equipe d'évaluation : Bruno Crépon (Crest-Insee et Jpal), Marc Gurgand (Paris School of Economics, Crest-Insee et Jpal), Luc Behaghel (Paris School of Economics, Crest-Insee et Jpal)

Période de mise en œuvre : Dec 2007 juin 2008

Remise du rapport final d'évaluation : septembre 2009

Présentation de l'intervention : Les prestataires sélectionnés doivent mettre en œuvre un accompagnement renforcé reposant sur un suivi hebdomadaire et des rencontres régulières avec un référent unique. La prestation est adaptée à chaque demandeur d'emploi et comporte un suivi post-reclassement de 3 à 6 mois en fonction du public visé. Parallèlement à ces expérimentations de recours à des opérateurs privés de placement, l'ANPE (désormais Pôle emploi) a développé, à partir du 1er janvier 2007, une prestation expérimentale d'accompagnement vers l'emploi « Cap Vers l'Entreprise » (CVE), dans six régions. Cette prestation, d'une durée de six mois éventuellement prolongée de trois mois, est mise en œuvre par des équipes de l'ANPE.

Résultats : A l'horizon des 12 mois d'observation, les accompagnements renforcés mis en œuvre par les OPP et les équipes CVE ont un impact statistiquement significatif sur le taux de sortie des listes de l'ANPE, par rapport à l'accompagnement classique. Au bout de 12 mois, le programme OPP permet d'accroître le taux de sortie vers l'emploi de +5,6 points. Le programme CVE a permis de faire passer le taux de sortie vers l'emploi des demandeurs d'emploi qui en ont bénéficié de 39% à 46,3% (environ +7 points). A des horizons plus courts, les impacts sont plus contrastés selon les programmes et les populations.

Généralisation ou pérennisation de l'expérimentation : oui partiellement

1.1.1.4. Programme de prévention des ruptures dans l'apprentissage (Tulle) (AP Diieses 09)

Financier de l'évaluation : Haut Commissariat aux solidarités actives, Appel à Projets Diieses, puis complément Dares

Porteur du projet : Mission Locale de Tulle

Equipe d'évaluation : Bruno Crépon (Crest-Insee et Jpal), Marc Gurgand (Paris School of Economics, Crest-Insee et Jpal), Roland Rathelot (Dares Crest-Insee), Philippe Zamora (Dares et Crest-Insee), Thomas Le Barbanchon (Dares Crest-Insee)

Période de mise en œuvre : avril 2008 - dec 2011

Remise du rapport final d'évaluation : novembre 2012

Présentation de l'intervention : L'expérimentation a pour objectif d'accroître le nombre de contrats d'apprentissages signés en facilitant la rencontre entre offre et demande de travail, sur un marché marqué par la multitude et l'éparpillement des acteurs (Chambre des métiers, CFA, employeurs, Education nationale). Il s'agit également de prévenir l'échec en apprentissage en améliorant la qualité des projets des jeunes et des appariements entre jeune et employeur, ainsi qu'en organisant un système de veille pour repérer les difficultés qui peuvent survenir au cours du contrat. Pour ce faire, le dispositif mobilise, en amont de l'entrée en apprentissage, les conseillers de CFA et des missions locales pour accompagner les jeunes, les aider à construire leur projet et à rechercher un maître de stage. Il prévoit également un suivi tout au long du contrat d'apprentissage et notamment la mise en place d'une médiation en cas de tensions entre l'apprenti et l'employeur.

Résultats : Les différences de traitement entre le groupe test et le groupe témoin sont assez faibles. De ce fait, le manque de puissance statistique rend difficile la quantification précise des effets de l'accompagnement. Le programme d'accompagnement de la phase 1, à la recherche d'un employeur, ne s'est pas traduit par un accès accru à l'apprentissage. Néanmoins, il a favorisé un recours plus important au pré-apprentissage pour les plus jeunes. L'accompagnement en phase 2 a favorisé un turn-over des jeunes plus important, mais n'a pas réduit le taux d'abandon. L'accompagnement a semble-t-il permis à davantage de ruptures "latentes" de s'exprimer". L'évaluation n'a pas bénéficié d'une puissance statistique pour évaluer avec suffisamment de précision l'impact recherché.

Généralisation ou pérennisation de l'expérimentation : oui

1.1.1.5. Mallette des parents 6^{ème} (AP Diïeses 11)

Financier de l'évaluation : Haut Commissariat aux solidarités actives, Appel à Projets Diïeses

Porteur du projet : Rectorat de Créteil

Equipe d'évaluation : Marc Gurgand (Paris School of Economics, Crest-Insee et Jpal), Eric Maurin (EHESS, PSE, J-pal) , Nina Guyon (PSE, J-pal), Francesco Avvisati (EHESS, PSE)

Période de mise en œuvre : juil 2008 - juil 2010

Remise du rapport final d'évaluation : décembre 2011

Présentation de l'intervention : Ce programme consiste en trois réunions débats réunissant des parents d'élèves de sixième et des membres des équipes éducatives. Elles sont axées sur l'aide que les parents peuvent apporter aux enfants, les relations avec le collège et la compréhension de son fonctionnement. Une expérimentation

complémentaire a été menée l'année suivante (2009-2010) pour déterminer quelles méthodes étaient les plus efficaces pour augmenter le nombre de parents impliqués.

Résultats : Le premier effet de ce programme est un surcroît d'implication des parents volontaires, auprès de l'institution scolaire et une plus forte implication auprès de leurs enfants à la maison. Les parents des classes bénéficiaires ont davantage rencontré les enseignants (30% au lieu de 24% ont pris plusieurs rendez-vous), davantage participé aux activités des associations de parents d'élèves (35% contre 24%) ou estiment plus souvent avoir une bonne connaissance des options proposées (85% contre 76%). Les familles des classes tests ont été moins souvent convoquées par l'administration que celles des classes témoins. On note aussi une amélioration du comportement des enfants : moins d'absentéisme, moins d'exclusions temporaires, moins d'avertissements en conseil de classe, plus grande fréquence des distinctions lors du conseil de classe. Cette amélioration des comportements est également perceptible chez les enfants des parents non volontaires, qui n'ont donc pas participé aux débats, mais relevant des classes bénéficiaires.

Généralisation ou pérennisation de l'expérimentation : oui

1.1.1.6. Orientation des jeunes au lycée via des dispositifs de parrainage (Actenses) (AP Diïeses 01)

Financier de l'évaluation : Haut Commissariat aux solidarités actives, Appel à Projets Diïeses

Porteur du projet : Association Actenses

Equipe d'évaluation : Marc Gurgand (Paris School of Economics, Crest-Insee et Jpal), Luc Behaghel (Paris School of Economics, Crest-Insee et Jpal), Vera Chiodi (PSE J-pal)

Période de mise en œuvre : sept 2008 – juin 2011

Remise du rapport final d'évaluation : janvier 2013

Présentation de l'intervention : Le porteur de projet a mis en place un programme de parrainage afin d'accompagner les lycéens bénéficiaires du dispositif dans leur orientation scolaire et la formation de leur projet professionnel, en posant l'hypothèse selon laquelle le parrainage d'un jeune par un actif occupé peut affecter l'information et les aspirations des lycéens qui s'autocensureraient, afin d'élargir leurs choix d'orientation. Le dispositif propose aux élèves la mise en relation avec le parrain, des ateliers et des interventions conçus par l'association au cours desquels les parrains sont invités à participer, des sorties telles que des visites d'entreprises ou des salons liés à l'orientation ; les parrains peuvent quant à eux bénéficier d'une formation pour mieux accompagner les lycéens.

Résultats : Au total, l'intervention mesurée à l'échelle de la classe paraît assez peu intense, et aucun effet majeur et facilement interprétable sur l'orientation des élèves n'est identifié. Les élèves qui en ont bénéficié ont des ambitions plus modestes, ils sont sensiblement moins nombreux à désirer poursuivre leurs études en classes préparatoires, en BTS ou en IUT.

Généralisation ou pérennisation de l'expérimentation : oui partiellement

1.1.1.7. Internat d'excellence de Sourdun (HAP 1)

Financeur de l'évaluation : Haut Commissariat à la Jeunesse, Financement hors appel à projets

Porteur du projet : Rectorat de Créteil

Equipe d'évaluation : Marc Gurgand (Paris School of Economics, Crest-Insee et Jpal), Luc Behaghel (Paris School of Economics, Crest-Insee et Jpal), Clément de Chaisemartin (Crest PSE)

Période de mise en œuvre : août 2009 – dec 2012

Remise du rapport final d'évaluation : mai 2013

Présentation de l'intervention : Les internats d'excellence visent à promouvoir la réussite d'élèves motivés, mais qui ne bénéficient pas d'un environnement social ou familial favorable pour développer leur potentiel. Le régime de l'internat doit permettre à ces jeunes de travailler dans de bonnes conditions et de bénéficier d'un encadrement adapté tout au long de la journée. L'internat de Sourdun est le premier à avoir ouvert ses portes, à la rentrée 2009. Il a été suivi par 44 autres.

Résultats : Aucune différence au bout d'un an, ni en français ni en mathématiques entre le groupe témoin et le groupe test. Au bout de deux ans, les élèves internes obtiennent des résultats significativement plus élevés au test de mathématiques. La différence est de 41% d'un écart-type du score, ce qui signifie qu'en moyenne, un élève qui était initialement classé 45/100 atteint grâce à l'internat le même niveau que celui qui était classé 30ème. Il y a une absence d'impact en français mais ceci peut être lié à la taille limitée de l'échantillon ne permettant pas de détecter de différences lorsqu'elles sont petites.

Généralisation ou pérennisation de l'expérimentation : oui dans un premier temps puis suppression.

1.1.1.8. Tous en Stage (AP1 464)

Financeur de l'évaluation : Haut Commissariat à la Jeunesse, 1^{er} appel à projets

Porteur du projet : Association Réussir moi aussi

Equipe d'évaluation : Yannick L'Horty (TEPP, Université Paris-Est Marne la Vallée), Pascale Petit (TEPP, Université d'Evry Val d'Essonne), Emmanuel Duget (TEPP, Université Paris-Est Créteil)

Période de mise en œuvre : sept 2009 – juin 2011

Remise du rapport final d'évaluation : mai 2012

Présentation de l'intervention : L'accompagnement dont bénéficient les élèves de troisième consiste en quatre demi-journées passées avec un coach spécialisé issu du monde de l'entreprise. Les trois premières séances se déroulent avant le stage. Elles ont pour objectif de faire émerger la représentation d'un stage idéal, de créer l'envie de le réaliser, de définir les moyens de le trouver et de vérifier que chaque jeune a obtenu le meilleur stage possible. Pendant toutes ces séances, le collégien tient à jour un dossier de recherche de stage et est immergé dans un groupe de recherche (avec deux groupes par classe). La dernière séance a lieu à l'issue du stage. Elle a pour thème l'élaboration

du rapport de stage et fournit l'occasion d'un nouvel échange dans lequel les jeunes se projettent dans l'avenir.

Résultats : Ce dispositif d'accompagnement n'a d'effet ni sur l'effectivité du stage, ni sur la qualité du stage, évaluée au travers de la satisfaction du jeune, alors qu'il s'agit des objectifs visés par le dispositif. En revanche, il influence les choix d'orientation des élèves, ce qui correspond à l'objectif du stage lui-même. Les élèves accompagnés refusent moins fréquemment l'orientation vers des filières courtes et professionnalisées.

Généralisation ou pérennisation de l'expérimentation : oui

1.1.1.9. Améliorer la Qualité des emplois exercés par les étudiants (AQ3E) (AP1 336)

Financeur de l'évaluation : Haut Commissariat à la Jeunesse, 1^{er} appel à projets

Porteur du projet : Université du Maine

Equipe d'évaluation : Jaketerina DMITRIJEVA (TEPP, Université Paris-est Marne la Vallée), Yannick L'horty (TEPP, Université Paris-est Marne la Vallée), Pascale Petit (TEPP université d'Evry Val d'Essone),

Période de mise en œuvre : Sept 2009 – juin 2012

Remise du rapport final d'évaluation : mars 2013

Présentation de l'intervention : L'expérimentation met en place une cellule de placement par l'Université, en partenariat avec le tissu local d'entreprises, les intermédiaires du marché du travail et les collectivités territoriales. La cellule recueille auprès des entreprises des offres d'emploi à temps partiel pouvant être occupés par des étudiants qui souhaitent travailler. Elle organise l'appariement entre ces offres et les étudiants en fonction de leurs contraintes. Une cellule spécifique, réalise cette collecte et met les offres à disposition des étudiants par le biais d'une plateforme extranet.

Résultats : L'expérimentation a accru les chances d'accès des étudiants à un travail salarié. Les emplois qui ont été pourvus consistent plus souvent dans des missions ponctuelles, contractualisées, exercées durant la semaine et en journée (plutôt que le soir et le week-end), avec des horaires compatibles avec l'emploi du temps universitaire et une conciliation beaucoup plus facile entre études et travail. Enfin il n'y a pas ou peu d'effet sur la réussite des étudiants. Cela signifie que l'effet négatif d'un accès plus élevé à un travail salarié, a été compensé par les changements dans la nature des emplois occupés par les étudiants.

Généralisation ou pérennisation de l'expérimentation : oui partiellement

1.1.1.10. Inscrire les contrats en alternance dans une logique de parcours sécurisé (Beaune) (AP1 263)

Financeur de l'évaluation : Haut Commissariat à la Jeunesse, 1^{er} appel à projets

Porteur du projet : Mission Locale de Beaune

Equipe d'évaluation : Chirstine Guegnard (Institut de recherche sur l'éducation IREDU), Claire Michot (Iredu), Jean Bourdon (Iredu)

Période de mise en œuvre : oct 2009-juil 2011

Remise du rapport final d'évaluation : mars 2012

Présentation de l'intervention : L'expérimentation organise un accompagnement des jeunes en alternances en parcours d'apprentissage. Le groupe test bénéficie d'un tutorat renforcé, et le groupe témoin d'une simple information ; et tous les acteurs (employeurs, chambres consulaires, centres de formation, Éducation nationale) participent activement au processus d'amélioration de l'orientation, de l'accès à l'emploi et de la sécurisation des contrats. Le tuteur externe a pour principales missions d'assurer un accompagnement extraprofessionnel de l'apprenti, d'assumer un rôle de médiation au sein de l'entreprise en cas de difficultés relationnelles et, d'éviter le décrochage du jeune. Le projet instaure une politique de prévention autour de l'alternance : détection, signalement et médiation rapides avec possibilité de transitions sécurisées par changement d'entreprise ou d'orientation.

Résultats : L'accompagnement individualisé effectué par les Missions locales auprès des apprentis et des employeurs apparaît comme une réelle plus-value : il s'est révélé efficace en diminuant significativement le nombre de ruptures de contrat d'apprentissage et de départs du secteur de formation.

Généralisation ou pérennisation de l'expérimentation : souhait de continuer mais problème de viabilité de la structure faute de financements

1.1.1.11. CV anonyme

Financeur de l'évaluation : Commissariat à la diversité et à l'égalité des chances, Pôle Emploi.

Porteur du projet : Sous traité à Pôle emploi

Equipe d'évaluation : Luc Behaghel (PSE, Inra, Crest et j-pal Europe) Bruno Crépon (Crest et j-pal Europe), Thomas le Barbanchon (Crest), En collaboration avec la sous direction des études de Pole Emploi

Période de mise en œuvre : nov 2009- nov 2010

Remise du rapport final d'évaluation : mars 2011

Présentation de l'intervention : Depuis 2006, la loi sur l'égalité des chances a rendu obligatoire pour les entreprises de plus de 50 salariés le recours au CV anonyme. Mais le décret d'application correspondant n'a pas été publié. L'objectif fixé à l'expérimentation est d'éclairer la faisabilité et l'efficacité du recours au CV anonyme dans la lutte contre les discriminations. Cette expérimentation organise la suppression du bloc « état civil » d'un CV (Nom, prénom, adresse, date de naissance) dans 8 départements auprès des entreprises de plus de 50 salariées volontaires.

Résultats : Le CV anonyme agit contre la tendance des recruteurs à sélectionner les candidats du même sexe ou du même âge qu'eux ; mais comme on trouve des recruteurs des deux sexes et des recruteurs plus ou moins jeunes, les phénomènes d'homophilie se compensent d'un recruteur à l'autre et l'anonymisation du CV n'améliore pas, en moyenne, les chances des femmes ni des seniors. Surtout, le CV

anonyme dégrade les perspectives des candidats issus de l'immigration ou résidant en ZUS-CUCS. La généralisation du CV anonyme ne semble donc pas se justifier.

Généralisation ou pérennisation de l'expérimentation : non

1.1.1.12. Rebond pour les étudiants décrocheurs de l'enseignement supérieur (AP1 317)

Financier de l'évaluation : Haut Commissariat à la Jeunesse, 1^{er} appel à projets

Porteur du projet : Association AFIJ Association pour Faciliter l'Insertion des Jeunes

Equipe d'évaluation : Philippe Zamora (Crest, Jpal), Pierre Cahuc (Crest), Bruno Crépon (Crest Jpal), Florent Fremigacci (Crest, Université d'Evry Val d'Essonne)

Période de mise en œuvre : dec 2009 – dec 2011

Remise du rapport final d'évaluation : juillet 2012

Présentation de l'intervention : L'expérimentation vise à faciliter l'entrée dans la vie active d'étudiants décrocheurs. Elle a testé, différentes combinaisons d'actions pour l'insertion des étudiants en échec. Après une phase commune de diagnostic et de validation du projet professionnel, ces programmes incluent trois composantes supplémentaires, dont une partie des jeunes seulement bénéficie, suite à un tirage au sort :

- un volet spécifiquement tourné vers les métiers en tension ;
- une phase de formalisation et de validation du projet professionnel à travers le parrainage par des professionnels en activité ;
- la signature d'un contrat d'engagement réciproque.

Résultats : Il apparaît que ce sont essentiellement les traitements contrat et tension, et surtout le croisement de ces deux traitements, qui ont un impact significatif. Le parrainage a un impact nul sur l'accès à l'emploi mais un impact négatif sur le moral et le sentiment de bien être des jeunes: il semble constituer une source de frustration, sans doute liée à la découverte de multiples contraintes d'un monde professionnel que les jeunes connaissent mal selon l'interprétation des évaluateurs. Ils suggèrent de continuer à mener des expérimentations similaires

Généralisation ou pérennisation de l'expérimentation : souhait de continuer mais disparition de la structure faute de financements

1.1.1.13. Accompagnement partenarial d'itinéraires personnalisés d'accès à la qualification (APIPAQ) (AP1 359)

Financier de l'évaluation : Haut Commissariat à la Jeunesse, 1^{er} appel à projets

Porteur du projet : Mission Locale de Marseille, et mission d'insertion de l'éducation nationale

Equipe d'évaluation : Céreq : Dominique Maillard, Virginie Mora, Pascale Rouaud, Michel Théry, Patrick Veneau, Tania Lehberger (collab.)

Période de mise en œuvre : Dec 2009- dec 2011

Remise du rapport final d'évaluation : mars 2012

Présentation de l'intervention : Le dispositif est conçu pour les jeunes non qualifiés en rupture scolaire ou présentant des risques sérieux de rupture. Il consiste à mettre en place, en dehors de l'école, des lieux d'accueil pour les jeunes et leurs parents. Le dispositif repose ainsi sur un partenariat renforcé entre rectorat et mission locale, et sur la création d'espaces d'accueil personnalisés (EAP) animés par un organisme de formation indépendant. Ces accueils sont établis sous la forme de lieux d'écoute faisant intervenir plusieurs acteurs, dont les missions locales et les missions d'insertion de l'Education nationale. Les EAP sont donc le lieu d'un dialogue avec les jeunes, et ont pour objectifs le maintien ou le retour en établissement scolaire, l'accès à une formation professionnelle, la signature d'un contrat en alternance, afin de réduire significativement les sorties sans qualification du système scolaire.

Résultats : Résultats quantitatifs issue de l'assignation aléatoire (un pendant qualitatif a aussi été mobilisé dans l'évaluation de ce projet) : 149 jeunes du groupe intervention (29%) ont effectivement rejoint un EAP. La comparaison des groupes ne met pas en évidence de différence statistiquement significative sur les aspects relatifs aux formations suivies, aux diplômes obtenus après la rupture ou sur la situation du jeune à la date de l'enquête. Deux ans après le décrochage scolaire, le retour ou le maintien en formation et l'obtention d'un diplôme restent dans les deux cas extrêmement limités. Les jeunes du groupe « intervention » commencent un peu plus fréquemment une formation que ceux du groupe « témoin » mais ils abandonnent aussi cette formation avant son terme avec une plus grande fréquence. En dehors des objectifs d'accès à la qualification, les deux groupes voient de semblables proportions de jeunes en emploi à la date de l'enquête (15%) Au regard des objectifs assignés, ce résultat n'incite donc pas à préconiser une quelconque généralisation de ce dispositif expérimental

Généralisation ou pérennisation de l'expérimentation : partiellement

1.1.1.14. 10 000 permis pour réussir »

Financeur de l'évaluation : Haut Commissariat à la Jeunesse

Porteur du projet : Plusieurs Missions Locales

Equipe d'évaluation : Yannick L'horty (TEPP Université Paris Est), Emmanuel Duguer (TEPP, Université Paris Est), Pascale Petit (TEPP, Université d'Evry Val d'Essone), Bénédicte Rouland (TEPP, Université du Maine), Yiyi Tao (TEPP Université Paris Est).

Période de mise en œuvre : Janv 2010- sep 2010

Remise du rapport final d'évaluation : novembre 2012

Présentation de l'intervention : Cette expérimentation nationale consiste à proposer à 10 000 jeunes en difficulté d'insertion une participation financière significative, de l'ordre de 1000 €, et un accompagnement personnalisé pour leur permettre de passer le permis de conduire. L'expérimentation a été réalisée sur plus d'une centaine de territoires en France, en partenariat avec des missions locales et des auto-écoles sociales.

Résultats : A court terme, les résultats soulignent l'ampleur des effets de Lock-in dans le passage du permis de conduire en France, c'est-à-dire une limitation par cette formation d'autres investissements dans la sphère professionnelle ou privée. A long terme, deux ans après l'entrée dans le dispositif, ces résultats sont contrebalancés par des effets positifs sur l'insertion professionnelle des jeunes qui disposent du permis.

Généralisation ou pérennisation de l'expérimentation : pas d'information à ce stade

1.1.1.15. AgiLoJe (AP1 269)

Financier de l'évaluation : Haut Commissariat à la Jeunesse, 1^{er} appel à projets

Porteur du projet : Missions Locales et Associations

Equipe d'évaluation : Crédoc, Léopold Gilles, Elodie Alberola, Audrey Capron, Christelle Gris, Pauline Jauneau, Florence Tith.

Période de mise en œuvre : Juil 2010 – nov 2011

Remise du rapport final d'évaluation : mai 2013

Présentation de l'intervention : L'expérimentation AgiLoJe (Agir pour le logement des jeunes) vise à favoriser l'accès au logement autonome des jeunes de 16 à 30 ans en insertion professionnelle. Il repose sur l'hypothèse qu'un accompagnement renforcé permet d'améliorer l'accès au logement des jeunes, et de favoriser leur insertion sociale et professionnelle. Ce projet a été mis en œuvre sur 7 territoires par différents réseaux associatifs intervenant sur la problématique du logement et plus particulièrement sur le logement des jeunes, mais également par des Missions locales. Le dispositif expérimental repose sur quatre axes d'intervention principaux : Accueil-Information-Orientation; Accompagnement à l'accès et au maintien dans le logement; Partenariat renforcé des acteurs de l'hébergement, du logement, de l'insertion, de la formation et de l'emploi ; Prospection et mobilisation de logements.

Résultats : l'évaluateur ne peut se prononcer sur l'effet propre du dispositif sur le parcours résidentiel et professionnel des jeunes. En revanche, l'expérimentation a permis un enrichissement des services de droit commun (services préexistants ouverts à tous) qui bénéficieront de fait à l'ensemble des jeunes à l'avenir, et un renforcement des partenariats locaux qui favorisera la mise en cohérence de l'action des différents acteurs du logement.

Généralisation ou pérennisation de l'expérimentation : partielle

Ce travail de présentation effectué nous pouvons maintenant passer à la présentation de nos résultats empiriques en procédant à une analyse croisée de ces cas pour les trois séquences d'interaction au cœur de notre étude de l'expérimentation : la phase de conception, la mise en œuvre et la production et prise en compte de résultats.

2. La négociation des intérêts de la relation clinique : la phase de design

Cette première partie restitue les résultats issus de notre recherche empirique sur les études de cas d'expérimentations évaluées par la méthode aléatoire. Nous nous intéressons ici à la phase précédant la mise en œuvre de l'expérimentation. Cette étape est déterminante pour le reste de l'expérimentation et révèle la diversité d'intérêts d'apprentissages ainsi que l'équilibre des forces au sein de l'expérimentation. La compréhension des enjeux de l'expérimentation nécessite un détour préliminaire par une étude de ses acteurs, leur *champ d'action stratégique* d'appartenance et leur place au sein de ce dernier. Ces éléments permettent de comprendre quelles raisons les ont

amenés à s'engager dans cette entreprise et ce qu'ils souhaitent en retirer. Ensuite, ce travail nous renseigne sur les modalités de rencontre des parties prenantes de cette *relation clinique* ainsi que la négociation de l'équilibre des forces dans cette dernière. Ces éléments de contextes contribuent conjointement à définir le design de chaque expérimentation et expliquent la diversité d'incarnations de cette institution, résultat de l'équilibre atteint entre les activités de *travail institutionnel* contradictoires qui la composent.

2.1. Les raisons d'être des expérimentations aléatoires, la recherche de l'amélioration des positions d'acteurs au sein de champs en crise

La méthode expérimentale par assignation aléatoire organise la rencontre de deux types d'acteurs appartenant à des champs d'action stratégiques distincts. On retrouve d'une part le groupe des scientifiques composés d'économistes appliqués issus de différents centres de recherche. Ces acteurs peuvent être répartis en différentes catégories selon la place qu'ils occupent dans le champ d'action stratégique de l'expertise d'État. Cette position détermine en partie les intérêts qu'ils poursuivent dans la conduite d'évaluation d'expérimentations par assignation aléatoire. D'autre part, cette recherche s'intéresse au groupe composé par les porteurs de projets expérimentaux. Ces derniers peuvent être aussi répartis en trois catégories selon leur positionnement dans le champ d'action stratégique de l'action publique, notamment au regard de la conditionnalité du financement de leur action par l'État. Comme pour les évaluateurs, ce positionnement contribue à expliquer l'équilibre des intérêts qui les amènent à s'engager dans l'entreprise expérimentale.

2.1.1. Les évaluateurs/chercheurs, typologie de profils et d'intérêts

Il paraît surprenant de commencer l'étude de l'expérimentation par celle de ses évaluateurs au lieu de ses porteurs de projet. Néanmoins, les développements précédents de ce travail ont déjà souligné le rôle déterminant que les évaluateurs des expérimentations ont joué dans la promotion de cette pratique, voire dans l'élaboration même des projets expérimentaux. Nous soutenons dans cette recherche que l'expérimentation s'est diffusée en France d'abord par l'initiative de chercheurs qui souhaitaient conduire des évaluations expérimentales par assignation aléatoire, plutôt que par l'initiative des porteurs de projets souhaitant conduire des projets innovants. L'offre d'évaluation a donc participé à créer la demande et justifie de débiter la restitution de notre recherche par l'étude de ces acteurs. Nous proposons dans un premier temps de revenir sur leurs caractéristiques communes ainsi que sur les éléments qui les distinguent. Nous présenterons ensuite les intérêts de chaque catégorie d'acteurs à s'engager dans cette entreprise.

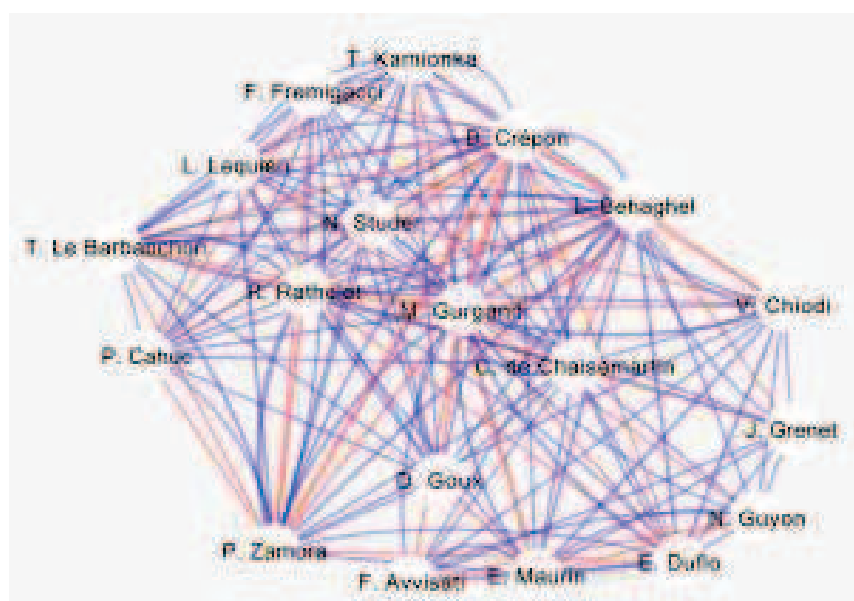
2.1.1.1. Les économistes appliqués, profil et typologie des chercheurs/évaluateurs

Les chercheurs/évaluateurs constituent un groupe aux programmes de recherches similaires, puisque tous sont des statisticiens qui étaient déjà associés à l'action publique et menaient des études quantitatives pour éclairer la décision publique. Aucun chercheur n'ayant auparavant participé à des études sur l'action publique n'est venu, grâce à cette méthode, s'associer à des acteurs publics. C'est bien parce qu'ils étaient déjà engagés dans des recherches appliquées qu'ils sont venus à la méthode aléatoire. Trois catégories

d'acteurs se dégagent de nos études de cas. Ces catégories sont conservées au cours de nos analyses et composent par la suite des facteurs explicatifs des variations d'intérêts et de relations avec le porteur de projet et le financeur.

D'abord, il est important de souligner que les évaluateurs scientifiques ayant recours à la méthode expérimentale sont peu nombreux. Seulement cinq centres ou fédérations de recherche se sont engagés dans cette entreprise et l'ont menée à terme. On retrouve, le Crest, L'École d'Économie de Paris, le TEPP⁸¹⁶, et de manière plus marginale le Céreq et le Crédoc. Une quarantaine d'évaluateurs ont participé aux expérimentations étudiées en composant des équipes de quatre chercheurs en moyenne. Néanmoins, certains chercheurs ont pu être associés à ces équipes sans pour autant pratiquer ces méthodes, notamment dans le cas de pluralité méthodologique comme pour le Céreq ou le Crédoc. Les chercheurs ayant participé à plus d'une expérimentation ne sont alors qu'une quinzaine. En dessinant les liens des équipes d'évaluation de chaque expérimentation et de leur appartenance de centre de recherche, on obtient une représentation en archipel où les associations de chercheurs se limitent aux appartenances de laboratoires, sauf pour le Crest et l'École d'Économie de Paris. Ces deux entités sont unifiées par un réseau dense d'associations inter-laboratoire, permis notamment par l'appartenance commune de ces chercheurs au J-PAL.

Figure 13: Représentation des liens entre évaluateurs par expérimentation (rouge) et par appartenance de laboratoire (bleu) de l'École d'Économie de Paris et du Crest



Lecture : Ce graphique a été construit grâce au logiciel d'analyse de réseau Cytoscape. Après avoir répertorié les chercheurs ayant participé aux expérimentations que nous étudions nous avons caractérisé en rouge les liens entre les chercheurs ayant participé aux mêmes expérimentations et en bleu les liens présentant les appartenances de laboratoire. Ce réseau comprend les chercheurs de l'Ecole d'Economie de Paris et du Crest. Il permet une représentation graphique des liens étroits entre ces deux laboratoires puisqu'ils apparaissent comme un même réseau dense.

⁸¹⁶ Fédération de recherche : Travail Emploi et Politique Publique

L'intégralité des chercheurs ayant participé aux expérimentations aléatoires a en commun de se définir comme « économistes appliqués ». Ils ont tous une formation en statistique ou en économie quantitative et ont tous préalablement conduit des travaux d'étude ou d'évaluation pour des commandes publiques, notamment la dizaine d'entre eux qui sont des administrateurs de l'Insee. Quant aux chercheurs du Céreq et du Crédoc, leur rattachement tutélaire aux ministères les inscrit traditionnellement dans des travaux appliqués à l'action publique⁸¹⁷. Le profil du praticien des expérimentations aléatoires est donc celui d'un chercheur de terrain, orienté vers l'action. Un administrateur de l'Insee ironise lors d'un entretien à propos de la méthode aléatoire : *« ce n'était pas du tout de la recherche fondamentale, c'est de la recherche hyper appliquée, ce n'est pas très compliqué d'ailleurs, au fond, c'est tellement appliqué que c'est abordable pour les Insee ! »*. Tous ces chercheurs étaient aussi engagés dans des travaux d'économie quantitative de modélisation ou quasi-expérimental, voire parfois de *testing* pour éclairer l'action publique⁸¹⁸. Un évaluateur du Crest explique sa démarche d'expert : *« J'adhère au discours qui dit que c'est important de cadrer l'action publique ou plutôt de guider l'action publique, j'ai la vision d'un foisonnement d'idées désordonnées, d'essais, de tentatives dont on n'a pas de retours, sans idée de la causalité de ce qu'on a vraiment fait. Je trouve ça assez intéressant d'essayer par des moyens expérimentaux ou non de produire de la connaissance quantitative ou qualitative sur la politique publique, je trouve que c'est primordial. »*

Malgré ces caractéristiques communes, notre analyse distingue trois catégories parmi ces acteurs. Ces catégories, basées principalement sur le statut de ces derniers dans l'expertise publique s'avèrent structurer aussi par la suite, leur intérêt et leur interaction avec le porteur de projet lors de l'expérimentation.

On retient d'abord, une catégorie composée de six acteurs centraux et expérimentés qui ont joué le rôle d'entrepreneurs promouvant cette méthode d'évaluation. Ces chercheurs ont environ une quarantaine d'années et bénéficient grâce à leur parcours au sein des Services Statistiques Ministériels d'un réseau important ainsi que d'une notoriété d'expert reconnue⁸¹⁹. Ils ont déjà mené de nombreux travaux d'études pour le ministère de l'Emploi ou de l'Éducation et sont des interlocuteurs privilégiés de l'État. Ils occupent des positions élevées, comme administrateurs hors classe, ou directeur de fédérations de recherche pour Y. L'horty par exemple. Ils ont chacun participé à au moins trois expérimentations sélectionnées pour notre étude. M. Gurgand et B. Crépon totalisent chacun, respectivement, 9 et 7 évaluations dans les cas que nous étudions. La plupart de ces chercheurs font partie du J-PAL Europe et se sont rapprochés de Martin Hirsch par leur participation à différentes instances scientifiques de conseil constituées par ce dernier. Certains ont participé, dès 2005, à la commission Famille, vulnérabilité, pauvreté, puis par la suite ont pris part au comité d'évaluation du RSA puis à la sélection des projets financés en 2008 par le HCSA (Haut Commissariat aux Solidarités Actives) et au Conseil Scientifique du FEJ dont M. Gurgand a été le président. Nous appellerons cette catégorie, les *titulaires* du champ d'action stratégique de l'économie

⁸¹⁷ Le Céreq, centre d'études et de recherche sur les qualifications, est un établissement public français à caractère administratif sous tutelle du ministère de l'Éducation et de ministre chargé de l'emploi. Le Crédoc, Centre de Recherche pour l'Étude et l'Observation des Conditions de Vie, est un organisme d'études et de recherche sous tutelle du ministre chargé de la consommation et du commerce. Il est subventionné à auteur environ de 10% de ses recettes par l'État.

⁸¹⁸ Le *testing* est une méthode expérimentale qui consiste à comparer la réaction d'un individu envers deux personnes ayant exactement le même profil pour toutes les caractéristiques pertinentes, à l'exception de celle que l'on soupçonne être la variable de discrimination.

⁸¹⁹ Les administrateurs de l'Insee ont vocation à travailler pour les différents services statistiques des ministères listés en annexe du Décret n° 2009-250 du 3 mars 2009 relatif à l'Autorité de la statistique publique, (2009) 2009-250.

appliquée qui sont caractérisés par des relations continues, directes avec et dans l'administration centrale⁸²⁰.

Une deuxième catégorie est composée d'acteurs moins expérimentés faisant partie des équipes d'expérimentations des chercheurs de la première catégorie. Ce sont de jeunes chercheurs qui sont soit en doctorat, soit au début de leur prise de poste et fortement liés au groupe de chercheurs précédent. À tout le moins, ils font partie des mêmes laboratoires, et certains conduisent même leur thèse sous leur direction. À titre d'exemple, E. Maurin est le directeur de thèse de N. Guyon et F. Avvissati, composant l'équipe de l'expérimentation « la mallette des parents ». Aussi, M. Gurgand dirige la thèse de L. Lequien, ils évaluent ensemble l'expérimentation sur l'accompagnement des RMIstes. Les acteurs composant ce groupe gravitent donc autour des acteurs *titulaires* et bénéficient de leur réseau, mais sans être les interlocuteurs directs auprès des décideurs publics. Ils sont davantage en interaction avec les acteurs de terrain lors des expérimentations. Leur rôle a pu évoluer au gré des expérimentations et certains de ces acteurs, au départ périphériques, ont pris plus d'importance au cours des années pour pratiquement rejoindre le statut des acteurs de la première catégorie, notamment grâce à l'accès à une affiliation J-PAL et à l'entretien des contacts qu'ils ont construit au fur et à mesure de leur carrière. Parmi ces acteurs périphériques, on retrouve aussi des chercheurs plus expérimentés qui s'associent ponctuellement à leurs collègues faisant partie du premier cercle. Chaque expérimentation menée par des acteurs de ce groupe nécessite la participation d'un acteur du premier groupe. Nous appellerons cette catégorie les *titulaires satellitaires* caractérisés par des relations continues et directes avec les *titulaires* qui les mettent en relation intermédiée avec l'administration centrale.

Enfin se dessine une dernière catégorie d'acteurs qui se distinguent par un statut différent, du fait de leur rattachement à des centres de recherche sous tutelle de l'État. Ces chercheurs font partie du Céreq ou du Crédoc. Ces centres de recherche bénéficient d'une part d'une subvention d'État pour conduire des travaux d'étude pour leurs ministères de tutelle comme les enquêtes-générations pour le Céreq par exemple, mais peuvent d'autre part, avoir une logique de « bureau d'étude », pour reprendre les termes de nos interlocuteurs, et répondre à des appels d'offres. Ce sont des acteurs scientifiques conduisant des travaux de recherche appliquée. Néanmoins, pour ces derniers la participation à des expérimentations constitue une activité nouvelle les amenant à travailler avec de nouveaux commanditaires comme par exemple des associations, et les amenant à pratiquer des évaluations de projets. Ces équipes témoignent d'une pluridisciplinarité où se rencontrent sociologues et économistes ayant recours à de l'analyse de données qualitatives et quantitatives. Ces acteurs entretiennent des liens indirects avec Martin Hirsch. Le Crédoc bénéficie d'une part de la présence de M.-O. Simon (ancienne directrice du département d'évaluation des politiques sociales du Crédoc) dans le Cabinet de M. Hirsch, et d'autre part de la présence de M. Angotti, directeur du même département, au Conseil scientifique du FEJ. Le Céreq a lui aussi accès au Conseil scientifique du FEJ, par la présence au sein de ce dernier du directeur du Céreq, A. Lopez. Cette catégorie s'apparente davantage à des *challengeurs* au sens où ils n'ont pas des relations directes, personnelles, privilégiées avec l'administration centrale et ont suivi dans le cas de l'expérimentation aléatoire, la norme préalablement établie par les *titulaires*. Les motivations de ces trois groupes d'acteurs pour s'engager dans l'expérimentation aléatoire sont variables, principalement selon leur appartenance à l'une de ces catégories.

⁸²⁰ Nous rappelons que ce concept de N. Fligstein et D. McAdam désigne des acteurs qui ont une influence disproportionnée au sein du champ et dont les intérêts et vues se reflètent fortement dans l'organisation dominante du champ, Neil FLIGSTEIN et Doug MCADAM, *A theory of fields*, op. cit. p. 15.

2.1.1.2. Les motivations des acteurs du champ de l'économie appliquée

S'intéresser aux motivations des évaluateurs des expérimentations est une étape indispensable pour comprendre leur recherche d'apprentissage et leur intérêt dans la conduite de cette méthode. Néanmoins, plusieurs types d'apprentissages sont recherchés dans cette pratique et d'autres facteurs de motivations sont aussi présents. Les entretiens conduits auprès des évaluateurs révèlent plusieurs raisons à leur engagement dans ce processus coûteux en temps et en énergie. Ces raisons sont souvent combinées, et le poids de chacune varie selon nos interlocuteurs. Nous revenons dans un premier temps sur ces motivations et proposons ensuite d'appréhender leur combinaison particulière selon chaque groupe d'acteur identifié précédemment.

Les économistes ont choisi précautionneusement les expérimentations auxquelles ils participent afin qu'elles nourrissent leur travail de recherche ou qu'elles leur permettent de s'inscrire dans des débats académiques contemporains. Un évaluateur résume le positionnement général de tous les évaluateurs rencontrés: *« en tant que chercheur, il y a des domaines et des thématiques qui nous intéressent particulièrement donc s'il y a des demandes [venant de porteurs de projet] qui correspondent à ces thématiques on va y aller, sinon, on n'ira pas. On ne cherche pas le financement, on n'a pas de gain particulier. Sur les sujets qui nous intéressent particulièrement, on va monter des projets pour essayer de trouver des financements »*. Ces chercheurs peuvent en effet, dans certains cas, orienter une partie du projet pour tester des théories débattues en économie, voire être les concepteurs du projet comme c'est le cas pour l'expérimentation « AQ3E » qui a été conçue par l'équipe du TEPP⁸²¹.

Les expérimentations leur permettent aussi d'avoir accès à de nouvelles sources de données primaires par leur association avec des porteurs de projet, dans des domaines jusque-là peu renseignés. Un évaluateur de l'expérimentation « permis de conduire » explique : *« là on a beaucoup plus réfléchi en terme de recherche. Les permis de conduire, c'est un gros investissement pour notre équipe, c'est 3 ans, et on s'est dit qu'on pourrait avoir plusieurs articles de recherche sur des thèmes différents qui n'avaient jamais été examinés ni en France ni ailleurs »*. D'ailleurs au-delà de l'étude de l'impact d'un type d'intervention, les chercheurs voient dans l'entreprise expérimentale un potentiel à générer, pour reprendre les termes d'un évaluateur du Crest, des « produits dérivés ». En effet, le contact direct avec le terrain grâce aux expérimentations constitue une occasion unique de générer des données primaires sur des populations jusqu'alors peu suivies. Dans le cadre de la même expérimentation, l'évaluateur précise que *« l'opportunité était trop intéressante pour ne pas observer d'autres éléments. On a fabriqué une sorte d'outil d'information, d'équipement scientifique qui va nous permettre de produire de la connaissance sur des tas de domaines qui sont a priori tout simplement inexplorés aujourd'hui. Donc, dès lors qu'on fait l'investissement énorme de construire le microscope qui va nous permettre de produire de nouvelles informations, il serait dommage de ne pas prévoir, et de passer à côté des données »*. D'ailleurs dans le cas de l'expérimentation « Rebond », on remarque même que dans les rapports d'évaluation reprenant les questionnaires d'enquête, le texte introductif destiné aux enquêtés mentionne une enquête scientifique sans la restituer

⁸²¹ On lit dans le rapport d'évaluation : *« Ce projet a été initié et conçu par l'équipe d'évaluation qui l'a proposé à l'Université du Maine pour l'expérimenter »*. LABORATOIRE ERUDITE, *Améliorer la Qualité des Emplois Exercés par les Etudiants (AQ3E)*, Rapport d'évaluation, 2013, p. 10.

dans le cadre évaluatif⁸²². On retrouve cet intérêt académique dans les publications issues des expérimentations. L'emploi de la méthode aléatoire permet aux scientifiques de s'inscrire dans des débats internationaux de micro-économie. Par exemple, pour l'expérimentation sur l'accompagnement des RMistes, les chercheurs soulignent qu'il y a débat sur l'efficacité du conseil et de l'accompagnement dans la recherche d'emploi. Alors que plusieurs travaux européens non expérimentaux établissent l'efficacité de ce type de dispositif, une expérimentation aléatoire aux Pays-Bas contredit ce résultat⁸²³. De fait, les résultats d'une expérimentation aléatoire supplémentaire en France permettront de répondre à ce débat.

En effet, un deuxième intérêt académique réside dans la volonté des chercheurs d'expérimenter l'utilisation de la méthode aléatoire, jusque-là inemployée dans l'évaluation des politiques sociales en France, afin de s'inscrire dans les débats internationaux⁸²⁴. Un chercheur/évaluateur de l'École d'Économie de Paris explique : *« cette méthode est devenue la norme en économie appliquée (...) c'est devenu dans la profession quelque chose de valorisé. Si on veut des résultats crédibles, on attend que la preuve soit donnée par cette méthode là. La collectivité scientifique est plus exigeante, car on sait que c'est faisable et que plein de gens l'ont fait »*. En effet, comme nous l'avons précisé précédemment, ces chercheurs étaient déjà tous inscrits dans le champ de l'économie appliquée et pratiquaient des méthodes alternatives dont l'enjeu était de reconstituer le plus fidèlement possible des groupes témoins permettant de comparer l'effet étudié sur les bénéficiaires avec un scénario contrefactuel. Nos interlocuteurs justifient leur attrait pour la méthode expérimentale par l'opportunité d'établir une causalité incontestable. Des évaluateurs de l'École d'Économie de Paris et du Céreq illustrent cette tendance par leurs propos. Le premier précise : *« c'est quand même rare qu'en économie on arrive à mesurer ce qu'on veut mesurer. Souvent on mesure le truc d'à côté et on se dit qu'on n'est pas trop loin. Alors que là on construit le protocole expérimental pour mesurer ce qu'on veut, et on le fait de manière relativement simple »*. Le second ajoute *« les papiers économétriques sur l'évaluation sont tout le temps tirés par les cheveux (...) Là, on a eu un peu une annonce : « enfin on a en sciences sociales une méthode incontestable, neutre, absolue !... »*. La pratique de la méthode aléatoire devient donc de plus en plus indispensable pour se positionner dans les débats d'économie appliquée internationaux. Un membre du cabinet de Martin Hirsch souligne aussi cette tendance : *« chez certains évaluateurs on avait le sentiment que ce qui primait c'était l'application de la méthode, alors que l'esprit de ce type d'appel à projets c'est quand même de ne pas être dans un pur exercice scientifique, mais de faire avancer les politiques sociales »*.

L'intérêt d'apprentissage de ces chercheurs est double, ils cherchent donc d'une part à générer des données sur l'impact d'une intervention et d'autre part à acquérir une expertise dans la conduite de la méthode aléatoire. Néanmoins, la particularité de leur démarche vis-à-vis des autres évaluateurs est que ces chercheurs ne se perçoivent pas comme des prestataires d'évaluation, mais avant tout comme des scientifiques pouvant aider à résoudre des questions économiques liées à l'intervention publique. Ils sont

⁸²² En témoigne une partie de ce texte introductif de questionnaire : *« Le Centre de Recherche en Économie et Statistiques (CREST) organise une enquête sur les jeunes qui s'inscrivent à l'AFIJ. Cette enquête porte sur les attentes, les projets professionnels et l'insertion future des jeunes qui sortent de l'Université. Ainsi le CREST vous appellera dans six mois et dans un an pour connaître mieux votre situation. Ces enquêtes ne seront utilisées que dans un objectif statistique. Vos réponses seront traitées de façon anonyme et totalement confidentielle »*. CREST, Rebond: programme d'accompagnement des étudiants décrocheurs de l'enseignement supérieur, Rapport d'évaluation, Fonds d'expérimentation pour la jeunesse, 2012, p. 52.

⁸²³ Bruno CRÉPON, Marc GURGAND, Thierry KAMIONKA et Laurent LEQUIEN, *Is Counseling Welfare Recipients Cost-Effective? Lessons from a Random Experiment*.

⁸²⁴ Pour un développement détaillé du développement de cette appétence, chez les statisticiens français se référer à notre 2^{ème} chapitre.

guidés d'abord par leur intérêt d'apprentissage général de connaissances nouvelles sur les mécanismes de l'action publique. C'est cette poursuite d'un agenda de recherche dans l'expérimentation qui nous amène à caractériser cette entreprise de *relation clinique*.

Enfin, cette méthode présente une combinaison pratique entre un travail de recherche et la volonté des chercheurs de pratiquer l'économie appliquée. Cette méthode a l'avantage de pouvoir être utilisée pour l'évaluation en tant que telle, tout en pouvant être convertie en production scientifique correspondant aux standards des revues scientifiques internationales. Ainsi, la pratique de la méthode aléatoire appliquée à des problématiques d'action publique permet l'apport de financements externes pour les équipes de chercheurs afin de financer les frais d'une recherche particulièrement coûteuse en enquête de suivi des bénéficiaires. Ces fonds ne vont pas directement aux chercheurs qui, pour reprendre leurs termes, « *n'en avaient pas besoin pour manger* », mais constituent une source de financement pour les laboratoires.

De plus, la méthode expérimentale est très coûteuse à conduire de par les enquêtes de suivi qu'elle nécessite et, sans financements conséquents, les laboratoires n'ont pas les fonds suffisants pour la mener. À titre d'exemple, l'évaluation « Jeunes Diplômés » de Pôle emploi a coûté environ 600 000 euros et l'évaluation de l'internat de Sourdun environ 930 000 euros⁸²⁵.

Ces trois intérêts, pour l'objet de l'expérimentation (et les données acquises), pour la méthode en elle-même (et sa maîtrise, valorisée dans le milieu économiste international) et pour les financements qu'elle apporte, sont présents chez tous les évaluateurs l'ayant pratiquée, néanmoins leur poids est différent selon leur catégorie d'appartenance.

Les *titulaires* cherchent avant tout à pratiquer cette méthode et à l'instaurer comme norme scientifique. Un économiste du CREST explique : « *le Graal des économètres c'était de résoudre les problèmes de sélection, et l'expérimentation était l'outil phare* ». Plusieurs articles témoignent de l'engouement de ces chercheurs pour cette méthode, portée à l'époque par la médiatisation des travaux d'Esther Duflor⁸²⁶. Un des évaluateurs de la « Mallette des parents » résume leur démarche pour aborder le projet : « *on voulait que ce soit une évaluation faite avec la méthode de l'assignation aléatoire* » ou, pour l'expérimentation sur l'accompagnement des RMistes, il ajoute : « *c'était vraiment très nouveau à ce moment-là, on voulait maintenir le coût le plus faible possible puisqu'on voulait vraiment apporter la preuve que c'était possible de faire ça* ». Grâce à leur positionnement stratégique en experts d'économie appliquée et leur proximité avec le cabinet de Martin Hirsch et les services statistiques ministériels, ils ont contribué à établir cette méthode comme favorite pour la sélection des appels à projets et la conduite d'expérimentation. *Titulaires* du champ d'action stratégique de l'expérimentation sociale, ils établissent cette méthode comme ressource clé pour accéder au statut d'évaluateur titulaire. Si cette démarche est d'abord issue de dynamiques personnelles, elles sont devenues peu à peu des dynamiques institutionnelles notamment avec la création de la fédération de recherche de l'École d'Économie de Paris et le rattachement

⁸²⁵ Peu de données sont publiées sur les financements accordés aux évaluateurs et ces derniers étaient assez réticents à nous communiquer ce montant. Le montant du financement de Jeunes Diplômés est issu d'entretien et celui de l'internat de Sourdun de COUR DES COMPTES, *Rapport public annuel, Des internats d'excellence à ceux de la réussite: la conduite chaotique d'une politique éducative et sociale*, février 2014, p. 196.

⁸²⁶ Esther DUFLO, « Evaluer l'impact des programmes d'aide au développement: le rôle des évaluations par assignation aléatoire », *op. cit.*; Yannick L'HORTY et Pascale PETIT, « L'évaluation aléatoire: un succès qui ne doit rien au hasard », *op. cit.*; Philippe ZAMORA, « La méthode d'évaluation aléatoire: apports et limites », *op. cit.*; William PARIENTE, « Analyse d'impact : l'apport des évaluations aléatoires », *STATECO* 2008.103.

du J-PAL Europe en son sein, dont l'activité principale est l'organisation d'évaluations expérimentales par assignation aléatoire⁸²⁷. La dimension financière est présente pour ces derniers puisque les financements issus des expérimentations permettent de diversifier les sources de subventions des laboratoires, et aussi parce que la conduite d'expérimentation aléatoire nécessite une ingénierie de collecte de données qui ne peut être prise en charge uniquement par le laboratoire. Se présente donc enfin, l'opportunité d'essayer cette méthode qu'ils ne peuvent entreprendre d'eux-mêmes en France.

Les *titulaires satellitaires* n'ont pas influencé les préférences institutionnelles pour la méthode aléatoire, mais leur intérêt pour cette méthode fait partie d'un positionnement stratégique. Ces derniers, en tant que jeunes chercheurs, souhaitent se positionner sur des techniques émergentes en économie appliquée en France, lesquelles leur ont donné accès aux débats académiques internationaux tout en gardant un contact direct avec le terrain. Cette méthode leur permet de s'associer aux travaux d'économistes réputés ainsi que d'avoir accès à de nouvelles données qu'ils peuvent valoriser dans leurs productions de recherche. La maîtrise de la méthode aléatoire, compétence encore rare en France, leur apporte donc un avantage non négligeable dans leur recherche de poste en économie appliquée, sur le territoire français ainsi qu'à l'étranger.

L'attrait pour la méthode aléatoire pour les *challengeurs* est d'une autre nature. Ces derniers n'ont pas un accès aussi privilégié aux unités internes de gouvernance du champ et n'ont donc pas participé en amont à l'établissement de cette norme. Néanmoins, ils ont assez de réseaux institutionnels pour être dans les réseaux d'information privilégiés des appels à projets, ils ont les compétences scientifiques pour conduire ce type de méthodes. Le choix de cette méthode est donc avant tout stratégique pour augmenter les chances de sélection de leur projet par les appels à projets du Haut Commissariat. Un évaluateur du Céreq explique les raisons de son choix pour la méthode aléatoire : *« on ne s'est pas posé la question parce que c'était la méthode préconisée par le Fonds d'expérimentation (...). Dès le départ on l'a posé comme une contrainte forte, on leur a dit [aux porteurs de projets] : « on va faire comme ça ». Il n'y avait pas vraiment de discussion possible puisqu'on pensait que, pour que ce soit accepté par les financeurs, il fallait en passer par là »*. Bien sûr les chercheurs de ces centres de recherches sont aussi intéressés par la conduite de cette méthode en tant que telle puisqu'elle fait aussi partie de la boîte à outils de l'économie appliquée et qu'ils commencent à la rencontrer dans les débats académiques d'économie appliquée⁸²⁸. Néanmoins, pour cette catégorie d'acteurs les motivations liées à des préoccupations propres à leur appartenance institutionnelle sont davantage présentes. Depuis la mise en place de la RGPP à partir de 2007, les dotations d'État en direction de ces centres sont de plus en plus réduites menaçant leur survie. Le Céreq et ses réseaux de centres associés perdent des dotations et opèrent un tournant institutionnel vers une diversification de leur source de financement. Un interlocuteur du Céreq nous précise *« le deal a été très clair, c'était « soit vous ramenez de l'argent, soit on vous ferme ». Et donc, comme un seul homme, le réseau [de centres associés] s'est porté volontaire »*. Ce mode de travail était nouveau pour ces centres de recherches qui, bien qu'habités à conduire des études pour l'action publique, n'avaient jamais encore collaboré sur des projets à courts termes avec des porteurs de projets.

⁸²⁷ Un économiste du Crest explique : *« ça s'est beaucoup fait autour de chercheurs qui croyaient beaucoup en ça et qui ont essayé de le développer, mais pas du tout autour de volonté de laboratoire de dire « on va consacrer tant de nos effectifs à cette méthode » »*.

⁸²⁸ L'évaluateur explique : *« ils disaient (les défenseurs de la méthode aléatoire) « non, non, non, vous n'arriverez jamais à ôter tous les biais puisque votre design d'enquête n'a pas été fait sur du random pur » donc on a voulu essayer à Beaune »*.

Ces centres de recherche se sont donc massivement portés candidats comme évaluateurs de projets. Le Crédoc par exemple s'est présenté comme évaluateur pour 23 projets. Au final 28 projets de Céreq et 18 du Crédoc ont été retenus⁸²⁹. Selon nos interlocuteurs, la participation aux expérimentations a rapporté au Céreq environ 2,8 millions d'Euros permettant de compenser en partie la perte de subvention. Contrairement aux deux autres catégories d'évaluateurs, les membres de ces centres de recherche ont proposé une diversité de méthode d'évaluations qui cherchent à établir un groupe de comparaison témoin, mais qui néanmoins n'ont pas toujours recours à la méthode aléatoire. Pour le Céreq, sur les 28 retenues, seules 3 évaluations proposent un protocole aléatoire, et 4 pour le Crédoc.

On observe donc un équilibre des intérêts différent selon la position des acteurs dans le champ de l'économie statistique appliquée :

- Les *titulaires* ont un intérêt élevé pour la méthode expérimentale, l'intérêt financier est présent, mais il a surtout pour but de trouver des financements pour conduire cette méthode ;
- Les *titulaires satellitaires* saisissent l'opportunité scientifique de participer aux expérimentations, ils sont moins préoccupés par les enjeux financiers qui sont résolus par le travail institutionnel des *titulaires* ;
- Les *challengeurs*, dont l'intérêt financier est premier pour la plupart, la conduite de l'expérimentation aléatoire pour sa dimension scientifique les intéresse, mais elle reste en partie un moyen pour avoir accès à une diversité de financements.

Si l'on reprend l'appareillage conceptuel de la *théorie générale des champs*, on peut établir que le *champ d'action stratégique* de l'économie appliquée en France connaissait une crise de légitimité vis-à-vis de la contestabilité de ses estimations d'effets, ainsi qu'une crise de financements publics. La méthode expérimentale par assignation aléatoire apparaît comme une solution à cette crise en permettant, tout en estimant rigoureusement les effets d'une intervention publique, de se positionner dans les débats internationaux d'économie appliquée et de diversifier les sources de financement des centres de recherche appliquée en augmentant la part émanant des marchés publics. Les titulaires ont effectué un travail institutionnel de construction d'un nouveau sous champ de l'économie appliquée par la constitution du champ de l'expérimentation sociale et en plaçant la maîtrise de la méthode aléatoire comme ressource privilégiée. À ce stade, les acteurs de la deuxième et de la troisième catégorie ont contribué à un travail institutionnel de maintien de cette norme qui leur est favorable. De leur côté, les porteurs de projets, aux profils très variables, ont aussi été motivés par un alliage d'intérêts d'apprentissage et de recherche de positionnement stratégique lié à une crise de financement dans leurs champs.

⁸²⁹ « Entre 2009 et 2012, le département Evaluation des Politiques Sociales (EPS) du CRÉDOC a évalué seize expérimentations en faveur des jeunes et le département Evaluation des Politiques Publiques en a évalué deux » Isa ALDEGHI, Pauline JAUNEAU et Lara MULLER, *Bilan méthodologique des évaluations conduites par le Crédoc dans le cadre du Fonds d'expérimentation pour la jeunesse*, 300, coll. Cahier de recherche, CRÉDOC, 2012, p. 13. Les rapports du Céreq actent uniquement 27 projets, mais nous y ajoutons le projet évalué par l'IREDU et porté par la mission locale de Beaune puisque les chercheurs sont membres d'un centre régional associé du Céreq.

2.1.2. Les porteurs de projets expérimentaux, typologie de profils et d'intérêts

Les porteurs de projets évalués par la méthode expérimentale par assignation aléatoire sont d'appartenance institutionnelle très variable. Ils sont néanmoins unis majoritairement par leur appartenance au niveau de l'exécution de l'action publique, et sont non décisionnaires. C'est d'ailleurs cette appartenance qui explique leur motivation d'engagement dans l'expérimentation. Bien que chaque porteur recherche un apprentissage sur son propre projet, il cherche avant tout par l'évaluation scientifique de son expérimentation à donner à voir l'effet de son projet afin d'en pérenniser le financement et de défendre le bien fondé de son intervention vis-à-vis d'un financeur et décideur public.

2.1.2.1. Les acteurs de l'action publique, profil des porteurs de projets

Les porteurs des expérimentations sont les acteurs de terrain de l'action publique. Ces derniers sont la cible de l'ouverture du champ de l'expérimentation sociale puisque le lancement d'appels à projets a pour vocation de faire émerger des « *initiatives innovantes des acteurs de terrain* » aptes à recevoir des subventions d'État⁸³⁰. Néanmoins, l'identité et les profils de ces derniers sont très variables. On compte des associations de petite taille très récentes comme Actenses fondée en 2006 (projet « Rebond ») ou « Réussir moi aussi » fondées en 2008 (projet « tous en stage ») ; des organismes du Service Public de l'Emploi, comme Pole Emploi et les Missions Locales qui organisent une mission de service public d'insertion auprès des publics de 16 à 25 ans⁸³¹ ; des Conseils Généraux, des Universités, des Rectorats et des directions générales de ministères. Ces porteurs de projet ont été associés en amont aux travaux de commission ou aux consultations organisées par Martin Hirsch, parfois dès la commission « famille, vulnérabilité, pauvreté », ou l'expérimentation du RSA, voire pendant l'élaboration du livre vert pour la jeunesse initié dans le cadre du FEJ.

Ces acteurs se caractérisent par une activité proche du terrain de l'action publique. Un acteur de Mission Locale explique « *on a une approche très opérationnelle des choses, on n'a pas une approche scientifique, car c'est un autre métier. Nous, on est là pour que les parcours se passent le mieux possible, à un coût raisonnable et en respectant les gens engagés dans ces parcours. Après...nous demander plus, c'est beaucoup* ». Peu de ces acteurs de l'action publique sont des personnels élus. Même dans le cas de projets portés par les Conseils Généraux, les évaluateurs rencontrés nous ont souligné que les personnes opérant les choix méthodologiques sont les personnels administratifs comme les directeurs généraux des services. Bien que dans l'opérationnel, certains acteurs peuvent être néanmoins politisés, notamment par leur nomination, comme les recteurs ou les directeurs généraux des services des ministères, numéro deux après le ministre, la pratique de l'expérimentation ne témoigne donc pas d'une diffusion de la pratique de l'évaluation chez les élus, pour qui l'évaluation électorale reste prédominante. Cependant, nombre de ces acteurs sont novices dans la commande d'évaluation externe. Ils connaissent surtout le renseignement d'indicateurs, comme de performance, de réalisations ou de résultats, avant tout par leurs outils internes de suivi, mais, mis à part Pole Emploi, aucun n'a déjà participé à la conduite d'évaluation externe s'intéressant à

⁸³⁰ Source : présentation du FEJ sur son site internet : <http://www.experimentation.jeunes.gouv.fr/1105-qui-sommes-nous.html>

⁸³¹ Code du travail, chapitre IV: Missions locales pour l'insertion professionnelle et sociale des jeunes, Article L5314.

l'impact de leur intervention de manière scientifique⁸³². Ce manque de culture d'évaluation scientifique n'est pourtant pas un obstacle au choix de la méthode aléatoire de par sa clarté de démonstration. Un évaluateur du Crest explique à propos de cette méthode : *« elle est facile à expliquer à des gens qui ne sont pas économistes, c'est facile de les convaincre que ce qu'on fait, ils peuvent comprendre vite et ils peuvent comprendre que ce que l'on mesure ça a un sens. Alors que si on leurs dit, « là il y a un grand modèle avec 43 équations empilées sur les autres et à la fin on trouve que ça fait 0,2 » alors les gens ne sont pas convaincus (...) De pouvoir leur expliquer ce qu'on allait faire avec les mains c'était un gros avantage de cette méthode. Avec les méthodes aléatoires on peut expliquer ce que l'on fait du début à la fin »*

Tous nos interlocuteurs ont souligné que les profils personnels ont joué un rôle important dans le choix des porteurs pour la méthode aléatoire. En effet, malgré leur diversité de profils institutionnels, on observe des caractéristiques communes aux porteurs de projets. D'abord, ces porteurs de projets sont animés par un intérêt scientifique. Ce dernier est variable, il peut aller d'une *« passion pour les sciences sociales »*, d'un *« esprit scientifique »*, à une dynamique de reprise d'étude, un parcours universitaire en statistique pour un acteur de la DGEFP voire au suivi d'une formation J-pal à l'expérimentation aléatoire pour des agents de Pole-Emploi qui conçoivent à cette occasion l'expérimentation du CV anonyme. Le porteur de l'expérimentation de « la mallette des parents » explique : *« j'ai été souvent perçu comme un original à vouloir faire bouger cette administration en créant des avant-gardes par l'évaluation et l'expérimentation. C'était ma marque de fabrique totalement revendiquée »*. Les évaluateurs sont parfois surpris de cet engouement scientifique, l'un d'eux déclare : *« le porteur a toujours été un partisan [de la méthode aléatoire], c'était son esprit scientifique ou autre, même s'il avait encore moins de scrupule et plus de volonté que nous »*. Cette appétence pour la science est incarnée particulièrement par le recteur de l'académie de Créteil qui initia trois expérimentations aléatoires. Il revient sur le recrutement de sa chargée de mission qui organisera la coordination des expérimentations : *« Je l'ai recruté au bout de quelques mois, j'ai vu une dépêche de l'AEF (Agence Education Formation) sur sa thèse qui venait d'être soutenue, sur la comparaison de l'éducation prioritaire en France et aux États-unis, avec une forte dimension d'évaluation des politiques publiques. Je me suis dit « formidable ! c'est une occasion de joindre science et action ». (...) Elle a su extrêmement bien faire des liens entre les théories qu'elle pratique, les acquis scientifiques et les choses très concrètes, c'est elle qui a fait le lien avec le cabinet de Martin Hirsch, et d'ailleurs, ultérieurement je recrutais à la DGSCO l'un des membres du cabinet de Martin Hirsch »*.

Il est possible d'identifier trois catégories de porteurs de projets selon leur positionnement dans le *champ d'action stratégique* de l'action publique, et notamment leur légitimité d'action, ainsi que la sécurisation de leur financement. On observe un premier groupe d'acteurs que l'on peut identifier comme les *titulaires* de ce champ. Ce groupe comprend les organismes financés principalement par l'État, détenant le monopole d'intervention dans leur domaine et dont la légitimité d'intervention est actée. On y retrouve par exemple les Rectorats, ou Pole Emploi (avant l'ouverture aux

⁸³² La pratique évaluative distingue dans l'intervention publique trois degrés d'action, les réalisations, les résultats et l'impact. Les réalisations désignent l'intervention en tant que telle par exemple la construction de pistes cyclables, les résultats désignent comment cette intervention a réussi à mobiliser les publics, l'emprunt de ces pistes cyclables et l'impact désigne l'effet souhaité par cette politique, comme par exemple la réduction de la pollution. Pour une discussion en détail de ces concepts voir Agathe DEVAUX-SPATARAKIS, « L'évaluation "basée sur la théorie" entre rigueur scientifique et contexte politique », (2014) 31-1 *Politiques et Management Public* 75-91.

opérateurs privés en 2005) ou les Conseils Généraux. Ceux-ci sont des *titulaires* du champ stratégique de l'action publique.

Un deuxième groupe d'acteurs est composé des organismes remplissant des missions d'intérêt public, mais dont les subventions d'États sont renégociées périodiquement. Ce sont des interlocuteurs privilégiés qui participent aux grandes concertations sur les nouvelles orientations de politiques publiques, mais dont les domaines d'intervention sont susceptibles d'être redéfinis, comme par exemple les grandes associations ou les fédérations⁸³³. Ces acteurs ne sont ni des *challengeurs* ni des *titulaires*, ce sont des *titulaires précaires* dont la position est renégociée périodiquement.

Enfin un troisième groupe est composé de nouvelles associations, de petites structures, ne bénéficiant pas de conventions de financements institutionnalisées avec l'État et dont l'intervention n'a pas encore acquis une légitimité. Ces derniers peuvent être identifiés comme des *challengeurs*, de nouveaux acteurs du champ de l'action publique.

Comme nous l'avons déjà développé précédemment pour les évaluateurs, ces différents profils de porteurs de projet constituent un facteur explicatif des raisons de leurs motivations à conduire des expérimentations.

2.1.2.2. Les motivations des acteurs de l'action publique

Pour les porteurs de projets, la décision de conduire une expérimentation évaluée par cette méthode représentait un double intérêt, l'un financier, l'autre de démonstration de la pertinence de l'intervention. Néanmoins, selon leur position dans le *champ d'action stratégique* de l'action publique, le poids respectif de ces deux motivations est variable

Avant tout, dans le cadre des expérimentations issues des appels à projets, les expérimentations constituent un apport de financement nouveau pour les porteurs de projets dans un contexte de restriction budgétaire. En outre, elle permet aux acteurs de terrain de s'extraire le temps d'un projet, d'une logique de prestation de service répondant à des commandes d'État et de porter des projets émergés du terrain de leur propre initiative. Le Fonds d'Expérimentation pour la Jeunesse finance à hauteur de 50% du projet expérimental et 100% de l'évaluation⁸³⁴. On retrouve deux logiques de financements, l'une consiste à profiter du financement pour organiser un nouveau dispositif comme ce fut le cas pour la « Mallette des parents », l'autre à donner une nouvelle ampleur et structurer des projets existants comme pour le cas de l'expérimentation conduite par la Mission Locale de Tulle. Cet apport est conséquent puisque, à titre d'exemple, cette subvention du FEJ ou du Haut Commissariat a atteint 620 669 euros pour la « Mallette des parents », 241 585 pour le projet « APIPAQ », 379 117 pour l'expérimentation de la Mission Locale de Beaune, 1 495 000 pour le projet « Rebond » et pouvaient atteindre jusqu'à 2,1 M€ pour l'internat de Sourdun⁸³⁵.

⁸³³ Pour un retour sur la relation financière précaire des Missions Locales voir le chapitre de Philippe Brégeon dans l'ouvrage Philippe BRÉGEON, *À quoi servent les professionnels de l'insertion?: l'exemple des intervenants sociaux en centres d'hébergement et de réinsertion sociale, en missions locales et dans l'insertion par l'activité économique*, coll. Le travail du social, Paris, L'Harmattan, 2008.

⁸³⁴ CONSEIL SCIENTIFIQUE DU FEJ, *Rapport du Conseil scientifique du Fonds d'expérimentations pour la jeunesse, pour la période mai2009-décembre 2010*, op. cit.

⁸³⁵ Source : entretien avec le porteur du projet, nous ne disposons pas des chiffres exacts pour le financement de la Mission Locale de Tulle et RECTORAT DE CRETEIL, *La Mallette des parents, APDIIESES-11, Note de restitution finale remise par le Rectorat de Créteil, porteur du projet, au Fonds d'Expérimentations pour la Jeunesse dans le cadre de l'appel à projets lancé en 2008 par le Ministère en charge de la Jeunesse*, Rapport d'évaluation, 2011; COUR DES COMPTES, *Rapport public annuel, Des internats d'excellence à ceux de la réussite: la conduite chaotique d'une politique éducative et sociale*, février 2014, p. 196; MISSION LOCALE DE BEAUNE, « *Inscrire les contrats en alternance dans une logique de parcours sécurisé* », *Note de restitution finale*, Fonds d'expérimentation pour la jeunesse, 2012; AFJJ, « *Rebond* », *Note de restitution finale du porteur de projet*, Mission d'animation du Fonds d'expérimentations pour la jeunesse, 2013.

Un porteur de projet résume : *« il y avait moyen d'avoir des interlocuteurs de plus haut niveau et d'aller plus loin dans la démarche, c'est ce qui nous a plu tout de suite et, il ne faut pas se leurrer, on était aussi en recherche de financements. À cette époque, il y avait un vrai coup d'arrêt des financements »*.

La dynamique de recherche de financements varie selon le profil des porteurs de projets. Elle est faible quand les porteurs ont des financements de l'État institutionnalisés comme dans le cas des expérimentations de Pole Emploi, et prédominantes pour les fédérations ou associations dont la survie dépend de subventions d'États. Comme pour les centres de recherches décrits précédemment, ces dernières connaissent des coupes de subventions importantes et sont à la recherche de nouvelles sources de financements. On retrouve notamment dans cette démarche les missions locales dont le CNML (Conseil National des Missions Locales) a organisé la dynamique de proposition de projets expérimentaux pour garantir leur financement⁸³⁶. Entre ces situations d'indépendances et de fortes dépendances financières, on retrouve un ensemble de porteurs de projets dont la survie ne dépend pas de ce type de financements, mais qui ne détient pas de fonds pour mener ce type d'initiatives. Par exemple, les Universités ou les académies ne disposent pas de fonds propres pour mener des projets innovants. Ces appels à projets leur donnent une marge de manœuvre nouvelle pour financer ce type de dispositifs. Un porteur de projet d'un rectorat explique : *« on avait des crédits que l'on n'aurait pas eus autrement. 95% du budget de l'éducation nationale sont des budgets de personnels et donc, quand il s'agit de faire des choses qui coûtent des sous, c'est impossible, et quand il y a une baisse du budget, elle se fait là dessus »*. Dans cette recherche de financements, les porteurs de projets cherchent dans leur appariement avec les évaluateurs de projets à augmenter leurs chances de sélection. Puisque le guide à l'intention des porteurs de projets présente la méthode aléatoire comme celle à favoriser, ils accueillent favorablement la proposition de ce type d'évaluation par l'évaluateur.

La deuxième motivation des acteurs réside dans leur volonté de démontrer la pertinence de leur intervention et de se légitimer comme acteur compétent. La démarche expérimentale a permis aux acteurs de donner à voir aux autorités compétentes le projet qu'ils souhaitaient mettre en œuvre, en y apposant une légitimité scientifique indépendante. Le porteur de l'expérimentation de Tulle précise : *« Quitte à trouver un évaluateur, autant trouver le plus sérieux qui soit, qui ne soit pas contestable (...) pour établir l'efficacité du projet »*. Cette perspective consistait à importer une pratique de généralisation scientifique dans l'environnement de la décision politique. Les porteurs de projets ont fait l'hypothèse que la méthode expérimentale leur permettrait d'établir l'effectivité de leur projet, et ainsi d'outrepasser les méandres réguliers de la mise à l'agenda de nouveaux types d'intervention ainsi que leur financement⁸³⁷. En effet, au-delà de la promotion de l'évaluation aléatoire, le FEJ s'inscrit dans une logique ascendante de l'élaboration de politiques publiques, et d'émergence de nouvelles solutions. Un acteur du FEJ explique : *« la logique du fond d'expérimentation, c'était de dire « laissons sa chance au terrain, qui dit avoir de bonnes idées, puisque là haut on n'arrive pas à résoudre tous les problèmes » »*. En s'appariant avec des évaluateurs scientifiques, les porteurs de projets cherchent à renforcer leur légitimité d'intervention sur de nouveaux champs d'action. Un évaluateur du Crest restitue les motivations d'un

⁸³⁶ Avec l'instauration de la LOLF, depuis 2008 chaque Mission Locales doit signer des conventions pluriannuelles d'objectifs (CPO) avec l'État, au sein desquels sont négociés des objectifs et des moyens susceptibles d'être renégociés périodiquement au vu de l'atteinte de résultats Source : <http://www.emploi.gouv.fr/acteurs/missions-locales>

⁸³⁷ La notion de *mise à l'agenda* en politique publique désigne « l'ensemble des processus qui conduisent des faits sociaux à acquérir un statut de « problème public » (...) justifiant une intervention publique légitime » Laurie BOUSSAGUET, Sophie JACQUOT et Pauline RAVINET, *Dictionnaire des politiques publiques*, op. cit..

porteur de projets : « *ils voulaient vraiment s'y inscrire* [dans les appels à projets] *à la fois pour des raisons d'argent et pour des raisons institutionnelles, pour se faire un nom* » ou, à propos des associations : « *ils croient beaucoup dans ce qu'ils font et ils vont se saisir du FEJ pour arriver à faire démontrer l'efficacité de leur dispositif* ». Il semblerait de plus, qu'au début du FEJ, les porteurs de projets ont cru comprendre que si l'évaluation de leur projet était positive, ils recevraient des financements de manière pérenne.

Reprenons les typologies établies précédemment. Pour les acteurs *titulaires* du champ de l'action publique, l'apport financier bien qu'important n'est pas la principale motivation, ils souhaitent avant tout légitimer leur action dans de nouveaux domaines d'intervention. Comme par exemple l'accompagnement des étudiants dans leur scolarité et la recherche d'emploi pour les Universités, ou la mobilisation des parents dans la scolarité de leurs enfants pour le rectorat de Créteil. Des acteurs *titulaires* peuvent aussi voir leur situation se précariser avec l'arrivée de nouveaux concurrents comme ce fut le cas pour Pole Emploi et l'ouverture à l'accompagnement des demandeurs d'emploi par les Opérateurs Privés de Placement. Dans ce cas, la méthode aléatoire est mobilisée pour apporter « *la démonstration flagrante* » de leur efficacité, pour reprendre les termes d'un interlocuteur de la Dares, de leur légitimité d'intervention et de leur savoir-faire, ce qui fut particulièrement le cas dans l'expérimentation « OPP/ CVE ».

Les *titulaires précaires* sont identifiés comme des acteurs légitimes dans leur domaine d'intervention, mais sont préoccupés par une sécurisation de leurs financements issus de subventions d'État. Les expérimentations sont souvent des formalisations d'actions déjà existantes et permettent de les financer et de sécuriser des postes. Ils cherchent surtout à démontrer leur pertinence d'intervention sur des domaines complémentaires de leur intervention régulière, et à établir de nouvelles gouvernances d'acteurs autour de leurs prérogatives.

Les *challengeurs* sont de nouveaux entrants dans le champ. Ils ne bénéficient donc pas pour l'instant ni de financements institutionnalisés ni de légitimité d'intervention reconnue. Ils sont donc enclins à accepter des contraintes fortes pour bénéficier de financements et se faire un nom dans le champ stratégique de l'action publique. On observe donc des équilibres variables d'intérêts selon la position de ces acteurs dans le champ de l'action publique :

- Les *titulaires* s'inscrivent principalement dans une démarche de légitimation d'un nouveau domaine d'intervention et la recherche de financements est là pour servir cette démonstration
- Les *titulaires précaires* sont dans un équilibre entre la recherche de renforcement de leur légitimité d'action ainsi que la sécurisation de leurs financements d'État.
- Les *challengeurs* sont ceux pour qui les enjeux de l'expérimentation sont les plus importants puisqu'ils sont motivés par la légitimation de leur action afin d'être reconnus comme acteurs du champ et par ailleurs l'expérimentation augmente considérablement leur apport de financements.

La recherche d'apprentissage par les porteurs de projets semble donc avant tout être orientée vers la communication envers un tiers, un financeur et décideur susceptible par la suite de pérenniser le financement et de généraliser le projet. Cette présentation de ces deux groupes d'acteurs – évaluateurs et porteurs de projets – et de leurs intérêts à conduire des expérimentations évaluées par la méthode aléatoires justifie de mobiliser le concept de *relation clinique* pour évoquer cette entreprise, puisqu'à l'origine aucune des

deux parties n'est au service l'une de l'autre et chacune poursuit son intérêt. De part et d'autre ces deux groupes d'acteurs cherchent à satisfaire plusieurs besoins grâce à l'expérimentation. Certains semblent converger, comme l'établissement d'un impact précis de l'intervention publique ainsi que la recherche de financements, mais d'autres apparaissent déjà comme potentiellement problématiques, notamment si l'impact estimé s'avère négatif, ou si la mise en œuvre du projet ne permet pas de mettre en œuvre de manière rigoureuse le protocole scientifique et empêche la production scientifique de données. Il est nécessaire à présent de revenir sur les contextes de rencontre de ces différents intérêts et de comprendre comment celle-ci a été encadrée par leur financeur.

2.2. La constitution de la relation clinique

La construction de la relation clinique expérimentale comporte deux temps, d'abord celui de la rencontre et de la décision de collaboration entre les deux parties, ensuite la négociation détaillée du design de l'expérimentation aléatoire. La rencontre peut être une « fausse » rencontre, au sens où plusieurs porteurs et évaluateurs se connaissent avant d'expérimenter ensemble. Elle peut être à leur initiative, par la mobilisation de leur réseau professionnel. Cette rencontre peut être aussi organisée par les financeurs et, dans ce cas, le cadre de cette organisation influence de manière importante l'équilibre des forces au sein de leur relation. La négociation autour de la conception de l'expérimentation est quant à elle favorable aux intérêts des chercheurs, et on observe par une comparaison diachronique, un apprentissage autour de cette phase de design qui prend de plus en plus d'importance pour déterminer le reste de l'expérimentation.

2.2.1. La rencontre des parties prenantes de l'expérimentation

Toute *relation clinique* commence par une rencontre. Les modalités de cette dernière ont une influence déterminante sur la suite de la relation. On retrouve plusieurs situations ayant abouti à la constitution d'une *relation clinique* ouvrant à la conduite d'une expérimentation aléatoire entre deux univers qui semblent de prime abord distincts. D'abord des rencontres constituées à l'initiative des parties prenantes sans intervention de la puissance publique. Les premiers cas d'emploi de la méthode aléatoire relèvent davantage d'une mobilisation de relations déjà existantes plutôt que d'une rencontre. On retrouve cette tendance particulièrement dans les cas d'expérimentations portés par Pole Emploi ou par des Universités. À défaut d'un réseau déjà existant, les parties prenantes des expérimentations se sont aussi sollicitées à l'initiative soit du porteur de projet, soit de l'évaluateur. Ensuite, avec la création du FEJ, le processus de rencontre a été de plus en plus institutionnalisé et contrôlé par les procédures de sélection des appels à projets. Les membres du Cabinet de Martin Hirsch ainsi que le service évaluation du FEJ endossent alors le rôle d'intermédiaire, appuyant la démarche des chercheurs dans l'équilibre des forces de cette relation.

2.2.1.1. Les rencontres issues de connexions interpersonnelles

La plupart des premières expérimentations aléatoires menées n'étaient pas l'occasion de nouvelles rencontres, mais plutôt le fruit de relations de proximité déjà existantes entre les scientifiques souhaitant pratiquer cette méthode et les potentiels porteurs de projets. Ces expérimentations sont principalement conduites par des chercheurs *titulaires* jouissant d'un réseau pénétrant les champs de l'expertise publique et du milieu scientifique. Les expérimentations issues de ces relations brouillent les grilles d'analyse

de la *relation clinique* puisque les acteurs y participant affichent des appartenances communes et des pratiques hybrides entre ces deux champs. Les commanditaires peuvent devenir prestataires d'évaluation et les évaluateurs devenir porteurs de projets. À ce stade ce ne sont pas deux mondes qui se sont rencontrés, mais un même monde à l'interstice de l'action publique et de l'économie appliquée qui organise des expérimentations. Les cas d'expérimentations aléatoires menées dans le domaine de l'emploi sont particulièrement emblématiques de cette démarche. On retrouve aussi cette tendance dans les projets expérimentaux portés par des universitaires.

Ces expérimentations, préalables aux appels à projets de Martin Hirsch sont le fruit de relations de confiance entre les chercheurs et les acteurs de l'action publique ou de l'expertise publique. Les chercheurs souhaitant conduire des expérimentations aléatoires ont bénéficié d'une relation privilégiée avec le personnel de la Dares et de la DGEFP⁸³⁸. Ces chercheurs avaient déjà effectué des travaux d'expertise dans le domaine de l'emploi, notamment concernant les politiques d'accompagnement entre 2001 et 2003⁸³⁹. Ils entretenaient des relations de confiance avec des administrateurs de la Dares, dont notamment la Sous-directrice du suivi et de l'évaluation des politiques de l'emploi, ainsi que le chef du département de la formation professionnelle et de l'insertion des jeunes et, par ailleurs, avec un autre administrateur, le chef du service des politiques de l'emploi et de la formation professionnelle de la DGEFP qui devient en 2008 le DGEFP⁸⁴⁰. Cet ensemble d'acteurs issus de la même formation scientifique (Ensaé) souhaitait développer les évaluations par assignation aléatoire. Un interlocuteur de la Dares précise : *« je connaissais bien les équipes du Crest et finalement c'est très important, (...) il me semble, vu l'implication qu'a eu la Dares dans ces projets là. Ça a été rendu possible par le fait que je connaissais très bien Bruno Crépon parce qu'on avait déjà travaillé ensemble sur différents sujets »*. Un évaluateur renchérit : *« on avait déjà travaillé avec eux, pour faire les évaluations de programmes d'accompagnement et quand ils nous ont contactés on leur a dit « c'est ça qu'il faut faire ». Non mais il faut reconnaître que parfois les choses tiennent surtout à des gens (...). Parfois il y a des gens qui décident de s'engager dans des directions qui sont résolument modernes. Je connaissais un peu les gens du service statistique avec qui j'avais travaillé, je leur avais expliqué et c'est eux qui m'ont accompagné en interne. C'était bien. C'est eux qui ont trouvé aussi que c'était une bonne chose de faire ça, car eux aussi ils étaient confrontés à ce problème de l'absence d'évidence (...). Le travail de la Dares, c'est aussi de sélectionner et de favoriser l'adoption... raisonnée... d'une nouvelle méthode. Eux ils étaient parfaitement dans leur rôle, lorsqu'ils ont décidé eux aussi de se joindre à l'opération. Ils ont participé au comité de pilotage d'OPP-CVE là, et après ils ont essayé de lancer des évaluations eux-mêmes. Et en particulier celle-là : « Jeunes diplômés »*. En effet, dès 2006, la Dares décide de soutenir la promotion de la méthode par assignation aléatoire. Cette promotion se concrétise par le recrutement d'administrateurs de l'Insee, R. Rathelot en 2007 (affilié J-pal) et T. Le Barbanchon en 2008, à des postes voués à la conduite d'expérimentations aléatoires. Ils partagent alors leur temps de travail entre la Dares et le Crest (respectivement 80% et 20% environ du temps de travail pour chaque institution)⁸⁴¹. La Dares a donc servi d'intermédiaire entre

⁸³⁸ DGEFP: Délégation Générale à l'Emploi et à la Formation Professionnelle

⁸³⁹ Bruno CRÉPON, Muriel DEJEMPE et Marc GURGAND, « Counseling the Unemployed: Does It Lower Unemployment Duration and Recurrence? », *discussion paper series IZA* 2005.

⁸⁴⁰ Ces informations sont issues de nos entretiens avec les parties prenantes de ces expérimentations, un interlocuteur de la Dares explique : *« le chef de service est devenu DGEFP quelques années plus tard. Il connaissait un peu les acteurs y compris Bruno Crépon ils se sont croisés à la prévision, il était plutôt partant et il nous a laissé faire »*

⁸⁴¹ Source : entretien avec le personnel de la Dares : *« pour développer des travaux d'évaluation quantitative aléatoire nous avons recruté à la Dares des administrateurs de l'Insee dévolu à ce sujet »* R. Rathelot a occupé ce poste de septembre 2007 à août 2010 avant de rejoindre le Crest, et T. Le Barbanchon de juillet 2008 à août 2011 avant de rejoindre le Crest à son tour.

les chercheurs et la DGEFP pour développer la méthode aléatoire. Par exemple, un administrateur de la Dares explique le lancement de l'expérimentation « Jeunes Diplômés » : *« ils nous disent (DGEFP) « voilà, on va faire l'expérimentation. Il faudrait que vous l'évaluiez. » Et à ce moment-là, je me rappelle très bien qu'on a proposé que ce soit aléatoire, mais on avait l'impression que ça allait les choquer ou que c'était infaisable (...) et on a eu l'aide énorme de Bruno Crépon qui est arrivé, qui était au Crest, et Esther Duflo(...) : l'évaluation était complètement financée par la Dares »*. Ainsi, les premières expérimentations par assignation aléatoire témoignent de l'importance du positionnement des *titulaires*, de leurs relations et d'une communauté de pensée avec les milieux de l'expertise traditionnelle des ministères. Cette position a été déterminante aussi pour les expérimentations « OPP/CVE », et le « CV anonyme ».

On retrouve ensuite cette même dynamique de réseau dans le cadre des appels à projets du Haut Commissariat, notamment dans les projets portés par les Universités. Par exemple, dans le cadre de l'appel à projet DIIESES ou le premier appel à projet du FEJ, l'équipe d'Y. L'Horty évalue des projets portés par des membres de sa fédération de recherche du CNRS : d'une part avec le projet « Nouveau dispositif d'orientation en première année de licence » porté par l'université Paris-Est Marne-la-Vallée et, d'autre part « AQ3E » avec l'université du Maine⁸⁴². D'autres acteurs activent leur réseau associatif, comme dans le cas de l'expérimentation « orientation des jeunes au Lycée via des dispositifs de parrainage » de l'appel à projets Diieses, portée par l'association Actenses. Cette association compte parmi ses membres actifs Marc Ferracci, chercheur au Crest et faisant partie de l'association de recherche d'Y. L'horty par son rattachement au laboratoire LEMNA de l'université de Nantes. M. Ferracci avait à cette époque déjà collaboré académiquement avec Bruno Crépon, il était lui aussi un promoteur de la méthode aléatoire⁸⁴³.

Au-delà des alliances bâties sur des relations préexistantes, on note aussi des relations issues d'une démarche individuelle, non organisée par les financeurs, d'un porteur de projets vers un évaluateur ou l'inverse. Par exemple, pour l'expérimentation « APIPAQ », c'est le porteur de projets qui s'est rapproché « naturellement » du Céreq, le porteur de projets explique : *« c'est venu naturellement. Je connaissais le Céreq, j'ai beaucoup lu ce qu'ils ont produit, c'est un partenaire naturel pour nous accompagner, j'ai rencontré l'un ou l'autre à une conférence je leur ai posé la question et très vite on a organisé une première réunion »*.

Certains évaluateurs sont aussi allés démarcher directement des potentiels porteurs de projets, comme par exemple dans le cas de l'expérimentation « la mallette des parents ». Un des porteurs de projets se souvient *« il [le porteur du projet] avait été approché par Éric Maurin (...). La question d'Éric Maurin c'était : « l'accompagnement éducatif, ça coûte des millions, on ne sait pas ce qui est le plus efficace dans ces types d'activités,*

⁸⁴² L'évaluateur explique : *« je suis directeur d'une fédération de recherche qui est sur cette université et ça nous permet effectivement de faire un réseau qui nous permet de monter des projets avec des universités »*. La fédération de recherche TEPP Travail, Emploi et Politiques Publiques (TEPP, FR n°3435) réunit Le Centre d'Études des Politiques Économiques de l'université d'Evry, EPEE, Université d'Evry Val d'Essonne ; le Centre Pierre Naville, CPN, Université d'Evry Val d'Essonne ; le Centre de Recherche en Économie et Management, CREM, Universités de Caen Basse-Normandie et Université de Rennes 1 ; l'Équipe de Recherche sur les Marchés, l'Emploi et la Simulation, ERMES, Université Panthéon-Assas Paris II ; l'Équipe de Recherche sur l'Utilisation des Données Temporelles en Économie, ERUDITE, Université Paris-Est Créteil et Université Paris-Est Marne-la-Vallée ; le Groupe d'Analyse des Itinéraires et des Niveaux Salariaux, GAINS, Université du Maine et le Laboratoire d'Économie et de Management de Nantes-Atlantique, LEMNA, Université de Nantes.

⁸⁴³ Se référer aux travaux suivant pour noter le positionnement favorable de Marc Ferracci envers la méthode aléatoire ainsi que son association avec les chercheurs du Crest la défendant Marc FERRACCI, *Evaluer la formation professionnelle*, Les Presses de Sciences Po, coll. Sécuriser l'emploi, 2013; Marc FERRACCI et Florine MARTIN, « Evaluation par expérimentation aléatoire de dispositifs de reclassement innovants pour les demandeurs d'emploi », (2013) 3-135 *Travail et emploi* 5- 26; Bruno CREPON, Marc FERRACCI et Denis FOUGERE, *Training the Unemployed in France: How Does it Affect Unemployment Duration and Recurrence?*, IZA, 2007.

est-ce que vous seriez d'accord pour qu'on le fasse avec un processus d'expérimentation par randomisation ? » (...) Ca a commencé comme ça, c'est Éric Maurin qui demande à rencontrer Jean-Michel Blanquer sur la base de l'accompagnement éducatif, et ce dernier qui réoriente sur les parents ». On retrouve aussi cette pratique avec certains Conseils généraux, un évaluateur de l'École d'Économie de Paris explique : « C'était supply driven [l'offre qui encourage la demande]. On a vu qu'il y avait un département qui voulait mettre en œuvre un accompagnement des rmistes, un peu par hasard dans une brochure, et on s'est dit « tiens, on pourrait l'évaluer de façon randomisée », et on a pris contact avec le directeur général des services. Et ils ont été convaincus par ce qu'on voulait faire. On a dû bien tomber. Il y avait l'envie d'évaluer l'impact de cette politique et de le faire comme ça. En l'espèce, sur ce cas, c'est venu des chercheurs ».

Ainsi, les premières expérimentations aléatoires ont bénéficié de conditions de *relations cliniques* privilégiées entre porteurs de projets et évaluateurs. Les porteurs avaient déjà une sensibilité scientifique ou alors étaient accompagnés par des acteurs intermédiaires organisant l'interaction avec les chercheurs. Pour les rencontres de ce type, le travail de définition en amont du design de l'expérimentation a été beaucoup plus maîtrisé par les chercheurs qui soit étaient familiers de ce secteur d'intervention de l'action publique, soit ont bénéficié de davantage de temps d'interaction avec le porteur. Néanmoins, avec l'ouverture du *champ d'action stratégique* de l'expérimentation sociale par la conduite d'appels à projets de grande ampleur, le cercle d'acteurs des expérimentations s'est élargi et ces rencontres interpersonnelles ont fait place à des rencontres structurées par les organisateurs des appels à projets.

2.2.1.2. Les rencontres organisées : « le « Meetic » de l'expérimentation sociale »

Les scientifiques n'avaient pas tous un réseau leur permettant d'identifier un porteur de projets souhaitant conduire une expérimentation sur un sujet rencontrant leur intérêt académique. Réciproquement, la plupart des porteurs de projets ne connaissent pas d'évaluateurs, ou pas ceux correspondants aux standards souhaités par le financeur. Ainsi, Martin Hirsch et son Cabinet, organisateurs des appels à projets, ont cherché à contrôler la rencontre entre évaluateurs et porteurs de projets afin que les expérimentations correspondent à leurs standards de qualité. Ces derniers se posent en entremetteurs, permettant la constitution de *relations cliniques* allant même parfois jusqu'à leur appariement sans que les deux parties aient un contact préalable. Ces rencontres s'avèrent souvent devenir conflictuelles dans la mise en œuvre de l'expérimentation, puisque l'alliance peut être plus facilement perçue comme subie et non maîtrisée par les deux parties.

Le premier appel à projets, initié par le cabinet du Haut Commissariat aux solidarités actives contre la pauvreté en septembre 2007, ne requiert pas que des couples déjà constitués candidatent. Les porteurs de projets doivent soumettre une manifestation d'intention avant le 9 novembre 2007⁸⁴⁴. Mais bien que les critères d'appréciation ainsi que des indicateurs d'évaluation doivent être identifiés par le porteur, il n'est pas requis de se présenter avec un évaluateur, puisque le Haut Commissariat propose d'accompagner cette mise en contact avant le dépôt de candidature le 1^{er} janvier 2008⁸⁴⁵. Martin Hirsch revient sur cette démarche et explique qu'ils ont essayé d'organiser « le

⁸⁴⁴ Il est précisé qu'un dossier de candidature pourra être déposé sans manifestation d'intention. Néanmoins il ne sera accepté que s'il met en relation un porteur de projet et une équipe d'évaluation

⁸⁴⁵ LE HAUT COMMISSAIRE AUX SOLIDARITES ACTIVES CONTRE LA PAUVRETE, *Appel à projet d'expérimentation sociale 2008*, 2007, p. 3.

"Meetic" de l'expérimentation sociale, en disant : soit ce sont des chercheurs et des universitaires qui ont des protocoles à proposer et on essaiera de trouver des collectivités d'accord pour être terrain d'expérimentation ; soit on a des collectivités, des entreprises, des associations, des acteurs locaux qui ont des projets et, dans ce cas-là, on essaiera de trouver des chercheurs pour pouvoir les aider à les mettre dans un protocole qui rende les choses véritablement évaluables et dans des conditions dans lesquelles on peut en tirer des leçons⁸⁴⁶».

Cette mise en contact s'est incarnée par l'organisation des « Rencontres de l'expérimentation », conférence de lancement du Grenelle de l'insertion du vendredi 23 et samedi 24 novembre 2007⁸⁴⁷. Les porteurs de projets ayant soumis une manifestation d'intention sont invités à y participer afin de mieux comprendre la démarche expérimentale et notamment les exigences d'évaluation. La méthode d'évaluation par assignation aléatoire est présentée comme un pré-requis pour garantir la qualité de cette démarche. En témoigne la présence et les interventions d'Éric Maurin (EHESS), Marc Gurgand (EEP), Bruno Crépon (Crest), et William Pariente (JPAL Europe)⁸⁴⁸. Le 24 novembre est organisé un Forum « meetic » des projets expérimentaux, où les porteurs de projets potentiels, qu'ils soient scientifiques ou acteurs de l'action publique ou associative, et ont déposé une manifestation d'intention dans le cadre de l'appel à projets, présentent leurs travaux et échangent, en présence des membres du comité de sélection⁸⁴⁹. Un des membres de l'équipe porteurs de projets de Tulle précise : « si je n'y étais pas allé je n'aurais pas compris ce qu'était l'évaluation par sélection aléatoire, et je n'aurais pas trouvé d'évaluateur c'est évident. Notre projet a répondu aux clous, car je suis allée là-bas (...), j'ai compris qu'il y avait des écoles de statistiques et des évaluateurs certifiés entre guillemets par le Haut Commissariat et que notre projet ne pourrait être retenu que si on avait ce type d'évaluateur. Si on allait chercher l'évaluateur du coin vers chez nous, ce n'était même pas la peine ». Ce ressenti est conforté par un évaluateur du Crest présent à ces journées : « Martin Hirsch a dit que c'était la méthode reine. Il ne disait pas qu'il n'y avait que ça comme méthode, mais que les projets avec cette méthode seraient bien regardés ». La sélection a d'ailleurs été très forte pour cet appel à projets puisque sur 640 manifestations d'intentions 240 dossiers ont été examinés et finalement, le 30 janvier 2008, 37 projets ont été retenus⁸⁵⁰. Il n'existe pas de documents renseignant sur la composition du Jury de cet appel à projets, mais plusieurs *titulaires* promouvant la méthode aléatoire y ont participé, comme notamment E. Maurin et Y. L'horty, sous la présidence de F. Bourguignon⁸⁵¹. Tous les projets proposant une évaluation expérimentale ont été retenus, soit 5 sur 37. On compte parmi ces projets : le projet parrainage porté par l'association Actenses, la « Mallette des parents », l'école de la deuxième chance, le programme de prévention des ruptures dans l'apprentissage porté par la Mission Locale de Tulle, et le « dispositif d'orientation en première année de licence ». Tous sont évalués par des titulaires du *champ d'action stratégique* émergeant de l'expérimentation aléatoire, c'est-à-dire des membres de l'École d'Économie de Paris, du Crest/Dares ou du TEPP⁸⁵². Sur ces 5 projets, seuls celui de l'école de la deuxième chance et de la mission locale de Tulle sont réellement

⁸⁴⁶ DARES, *Colloque sur les expérimentations pour les politiques publiques de l'emploi et de la formation*, op. cit. p. 8.

⁸⁴⁷ Lancement du Grenelle de l'insertion, les rencontres de l'expérimentation Sociale, programme, 2007.

⁸⁴⁸ Voir le résumé de cette deuxième journée sur le site <http://www.experimentationssociales.fr/vendredi.php>

⁸⁴⁹ Voir le résumé de cette deuxième journée sur le site <http://www.experimentationssociales.fr/samedi.php>

⁸⁵⁰ Olivier NOBLECOURT et Frédéric VALLETOUX, op. cit., p. 55.

⁸⁵¹ Source : entretiens et témoignage dans les actes du Colloque : DARES, *Colloque sur les expérimentations pour les politiques publiques de l'emploi et de la formation*, Paris, op. cit. p. 137.

⁸⁵² Nous lions ici l'affiliation Crest/Dares puisque les membres de la Dares ayant participé à cette évaluation bénéficiaient d'un rattachement aux deux organisations. Le TEPP (Travail, Emploi, Politiques Publiques) est la fédération de recherche CNRS dirigée par Yannick L'horty.

de nouvelles rencontres organisées par l'appel à projets, les autres correspondent davantage aux relations interpersonnelles décrites plus haut.

La position de l'équipe de Martin Hirsch, intermédiaire entre chercheurs et acteurs de l'action publique, s'est renforcée et institutionnalisée avec la création du FEJ. Des observateurs internes à cette équipe expliquent avoir procédé de manière « expérimentale » afin d'identifier une procédure de sélection des porteurs de projets et évaluateurs permettant de garantir la qualité scientifique des relations cliniques. Au fur et à mesure des appels à projets, les procédures de sélection ont été renforcées.

Le premier appel à projets (AP1) se caractérise par une situation d'urgence et de manque de contrôle du Conseil Scientifique et de l'équipe de Martin Hirsch sur la qualité scientifique des protocoles méthodologiques d'évaluation. Cet appel à projets requiert un appariement évaluateur/porteur de projets avant le dépôt des dossiers⁸⁵³. Néanmoins, La majorité des porteurs de projet ne connaissent pas d'évaluateurs externes et se tournent vers le FEJ pour avoir des contacts. Un membre de la Mafej se souvient « *le téléphone explosait, les mails n'en parlons pas (...) ils nous appelaient en disant « moi je ne connais pas d'évaluateurs », je leur disais : « prenez le bottin, adressez-vous à l'université d'à côté », mais vous imaginez bien que les laboratoires de recherche ils voyaient débarquer une petite asso... sauf à avoir déjà des contacts ou travaillé avec...* »⁸⁵⁴. La Mafej émet alors un guide à l'intention des porteurs de projets, préconisant dès que possible le recours à la méthode aléatoire, et elle organise dans l'urgence un appel à candidatures pour évaluateurs afin de publier une liste indicative d'évaluateurs selon la thématique⁸⁵⁵. Les évaluateurs sur cette liste sont alors sur-sollicités, et s'instaure alors une relation asymétrique avec les porteurs de projets. Les deux mois disponibles pour monter le dossier poussent les porteurs de projets à se référer aux évaluateurs sur cette liste⁸⁵⁶. Ces derniers se retrouvent rapidement débordés d'appels et donc dans une position dominante, sachant que leur refus d'association avec un porteur signifie souvent l'incapacité de ce dernier à proposer un projet dans les délais. Un interlocuteur du Crédoc raconte le processus qui les a amenés à sélectionner 22 projets sur 45 propositions : « *ayant été beaucoup sollicités, on était obligés de faire une pré-sélection avant la Mafej, surtout vers la fin. On savait que si on disait non [aux porteurs de projets], ils ne pouvaient pas déposer leurs projets et je trouve que ce n'était pas à nous de faire ce choix à la place de la Mafej* »⁸⁵⁷. L'offre de praticiens de l'évaluation aléatoire est encore limitée aux *titulaires* du champ et seuls quelques nouveaux centres de recherche s'engagent dans cette démarche⁸⁵⁸. Le Conseil Scientifique ainsi que les membres du cabinet de Martin Hirsch déplorent le manque de contrôle sur la qualité des expérimentations et décident de cadrer davantage la sélection des appels à projets. Dès le deuxième appel à projets, les critères évaluatifs sont détaillés pour chaque thématique et, à partir de l'appel à projets « lutte contre le

⁸⁵³ FONDS D'EXPERIMENTATION POUR LA JEUNESSE, *Programme national d'expérimentations pour les jeunes, premier appel à projets 6 avril 2009, 7 mai 2009*.

⁸⁵⁴ Pour rappel la MAFEJ est la Mission d'animation du Fonds d'expérimentations pour la jeunesse

⁸⁵⁵ CONSEIL SCIENTIFIQUE DU FEJ, *Guide méthodologique pour l'évaluation des expérimentations sociales à l'intention des porteurs de projets*, 2009.

⁸⁵⁶ Publication de l'appel à projet le 6 avril 2009, la clôture initialement prévue pour le 31 mai 2009 est prolongée au 15 juin 2009

⁸⁵⁷ « *Le CRÉDOC a été sollicité par de nombreux porteurs de projets pour évaluer en tout 45 projets. Pour 22 de ces projets, le CRÉDOC ne s'est pas porté évaluateur, soit parce que le porteur de projets n'avait pas trouvé de partenaire pour cofinancer l'expérimentation avant la date limite de l'envoi, soit parce que le CRÉDOC a considéré que le projet ne remplissait pas les conditions suffisantes pour être évaluable (objectifs peu lisibles, différence insuffisamment marquée avec les programmes déjà existants sur le territoire), soit encore parce que les délais étaient trop courts pour pouvoir élaborer un protocole d'évaluation satisfaisant* ». Isa ALDEGHI, Pauline JAUNEAU et Lara MULLER, *Bilan méthodologique des évaluations conduites par le Crédoc dans le cadre du Fonds d'expérimentation pour la jeunesse*, 300, coll. Cahier de recherche, CRÉDOC, 2012, p. 21.

⁸⁵⁸ Au final seuls 23 projets sur 167 ont proposé cette méthode et peu sont arrivés à terme.

harcèlement en milieu scolaire » en 2011, les projets et les évaluateurs sont présélectionnés à part dans un premier temps et peuvent s'apparier par la suite uniquement entre eux⁸⁵⁹.

Le recours à la procédure d'appel à projets pour encourager l'expérimentation aléatoire a été peu à peu délaissé en faveur des expérimentations hors appels à projets, ces dernières permettant un plus grand contrôle des membres de la *relation clinique*. Celles-ci s'apparentent à des mariages arrangés où le FEJ sélectionne séparément les porteurs de projets et les évaluateurs. Alors que la procédure d'appels à projets aboutissait à la signature d'une convention tripartite entre le FEJ, l'évaluateur et le porteur de projets, ici sont signés deux conventions bipartites l'une entre le FEJ et les porteurs de projets, et l'autre entre le FEJ et l'évaluateur. Dans ce cadre, les porteurs de projets ne choisissent pas l'évaluateur ou la méthode d'évaluation pour leur projet, et cette dernière est principalement la méthode aléatoire conduite par des *titulaires* du champ d'action stratégique⁸⁶⁰. Le choix des prestataires d'évaluation est centralisé au niveau des instances de direction du FEJ et du Conseil Scientifique, et défie parfois les règles de la mise en concurrence, comme le souligne la Cour des comptes pour l'expérimentation de l'internat de Sourdun⁸⁶¹. Un interlocuteur du FEJ explique : « *dans le cadre des évaluations nationales, il faut [que les porteurs de projets] se soumettre à des évaluations proposées, (...) c'était beaucoup moins riche dans la rencontre par rapport aux autres projets où l'association venait d'en bas* ».

Cela signifie que les porteurs de projets rencontrent l'évaluateur et découvrent sa méthode d'évaluation une fois qu'ils sont déjà sélectionnés et doivent l'accepter pour accéder au financement. Un des évaluateurs de l'expérimentation « permis de conduire » explique : « *ça ne s'est pas très bien passé au départ, parce que les porteurs de projet n'ont pas été informés qu'il y aurait une évaluation aléatoire (...). Ils avaient été très contents d'avoir été retenus, et là on les soumettait au tirage au sort (...). L'information n'avait pas été suffisamment explicite, donc il a fallu refaire un travail avec eux pour dire l'intérêt de ce type de méthodes* ».

Ainsi, de nombreuses *relations cliniques* ont été organisées par l'intermédiaire de l'équipe du Haut Commissariat. Ce soutien a été considérable pour influencer l'instauration d'un protocole d'évaluation aléatoire ainsi que sa bonne mise en œuvre par la suite.

2.2.2. La négociation du design de l'expérimentation

Succède à la signature de la convention d'appariement de l'équipe de recherche et des porteurs de projet, la phase de négociation, dite de « cadrage » dans les pratiques évaluatives. Elle organise les détails de l'application de la méthode aléatoire au projet. Les candidatures se faisant dans l'urgence, des précisions des modalités de mises en œuvre du protocole sont toujours nécessaires avant d'entreprendre le projet. Le témoignage d'un évaluateur du Céreq reflète cette tendance : « [les porteurs de projets

⁸⁵⁹ MAFEJ, *Programme national d'expérimentations pour la jeunesse, deuxième appel à projets*, 9 octobre 2009.

⁸⁶⁰ Sur les 11 expérimentations élaborées sur la période de 2009 à mi 2010 (pendant la présence de Martin Hirsch au gouvernement), 8 font l'objet d'une évaluation aléatoire.

⁸⁶¹ On lit dans ce rapport : « Les modalités de la commande publique ont été particulières, sans mise en concurrence sur la base d'un cahier des charges et sans qu'une étude préalable ne fonde la décision de s'abstenir d'une consultation en raison, par exemple, du savoir-faire particulier du prestataire choisi. Les règles usuelles de prise de décision du fonds d'expérimentation pour la jeunesse n'ont pas été non plus respectées ». COUR DES COMPTES, *Rapport public annuel, Des internats d'excellence à ceux de la réussite: la conduite chaotique d'une politique éducative et sociale*, février 2014, p. 196.

nous disaient] *« vite, vite, il ne nous reste que trois jours pour déposer un projet, vite » et dans bien des cas on a travaillé jusqu'à 11 heures du soir sur un petit coin de table de cuisine en disant « bon ben demain c'est bon ». Comment voulez-vous que dans ces conditions les projets aient une qualité scientifique en béton armé au moment du dépôt du dossier? »*

La sélection du projet requiert donc par la suite une période de renégociation détaillée du protocole qui dure en moyenne de 3 à 6 mois selon les difficultés rencontrées pour que chacune des parties prenantes de l'expérimentation puisse satisfaire ses besoins de connaissance en cohérence avec ses pratiques de champ. Nos études de cas témoignent d'une diversité de design d'expérimentations aléatoire reflétant ces compromis. Nous proposons dans un premier temps une analyse des forces en présence constituant cette nouvelle interaction entre les chercheurs et les acteurs de l'action publique, pour ensuite revenir sur les apprentissages issus de cette phase de rencontre.

2.2.2.1. Un équilibre des forces favorables aux chercheurs

Cette phase de travail est nouvelle pour les chercheurs et évaluateurs. Un des chercheurs du Crest explique : *« le gros intérêt de la recherche non randomisée, c'est que tu n'as pas vraiment besoin de demander l'autorisation et d'avoir l'accord des gens qui mettent en place le projet pour le faire (...). Personne ne peut t'empêcher »*. Ici l'accord du porteur sur la nature de l'intervention et le protocole est nécessaire à la production des résultats scientifiques. La phase de la négociation du design est le premier théâtre des tensions liées à cette association. Néanmoins, à ce stade, l'évaluateur bénéficie d'une position dominante. Lors de cette phase, aucun des évaluateurs titulaires du champ d'action stratégique n'a renoncé au protocole aléatoire⁸⁶².

Les évaluateurs bénéficient d'une position favorable dans l'équilibre des forces grâce à leur soutien par le Haut Commissariat organisateur et commanditaire officiel des expérimentations. Ils font alors face à des porteurs de projets qui ont choisi cette méthode en méconnaissance de cause ; un évaluateur du Crest souligne : *« les porteurs n'ont pas conscience de ce que ça implique, surtout au début »*. D'ailleurs certains porteurs ne l'ont pas choisi et en ignorent les implications pratiques. Un observateur du FEJ reconnaît : *« ce que l'on a beaucoup vu aussi, c'est des acteurs de terrain subir un peu la randomisation, mais se dire que c'était le prix à payer pour démontrer qu'ils ont raison, pour avoir des financements. On leur donnait l'opportunité de prouver l'efficacité de ce en quoi ils croyaient (...), c'était un mal pour un bien »*. Pour la majorité des porteurs de projet, l'expérimentation est leur première association avec des évaluateurs externes scientifiques. Des scientifiques, non habitués à interagir avec les porteurs de projets à ce stade et à devoir négocier les modalités de leur protocole de recherche, peuvent réagir vivement. Un statisticien du TEPP explique : *« quand on est contacté pour des expérimentations, on leur dit : « ou bien on évalue comme ça, ou on n'évalue pas ! » On a suffisamment de recul sur des méthodes quantitatives pour dire « C'est à prendre ou à laisser ! » Et des gens du J-Pal sont dans cette logique là aussi »*.

Les intermédiaires qui ont contribué à mettre en contact les deux parties de la *relation clinique* ont d'ailleurs dû intervenir dans plusieurs cas lors de cette phase critique afin de sécuriser la conduite d'un protocole aléatoire. Ce fut le cas par exemple de

⁸⁶² L'abandon de la « Cagnotte » évaluée par l'école d'économie de Paris est intervenu à ce stade, mais uniquement à cause d'une campagne de presse médiatique à charge. Et l'école de la deuxième chance évaluée par l'école d'économie de Paris a été abandonnée suite à des problèmes de mise en œuvre et ne sont pas intervenus à cette phase là.

l'expérimentation « Créa jeunes » porté par l'Adie, un évaluateur explique : « *J'étais allé la voir, je lui avais expliqué, elle avait dit oui, elle avait dit que ça allait et puis ensuite elle s'est un petit peu rétractée. Donc là, il y avait eu une pression de la part du FEJ pour lui dire « vous savez, nous, ce n'est pas tellement le projet qui nous intéresse, ça nous intéresse bien sûr, mais qu'à la hauteur de l'évaluation, c'est ça la règle du jeu. Sinon : pas d'évaluation, pas de projet. Donc du coup ils y sont allés »*. La Dares aussi a pu jouer ce rôle de facilitateur pour les évaluateurs, un administrateur explique pour l'expérimentation « OPP-CVE » : « *à un moment, on se demandait si ça allait se faire et finalement on a poussé pour s'introduire comme 3^{ème} acteur dans le sujet, avec le souhait de sécuriser l'opération* » ou pour l'évaluation « Jeunes Diplômés », « *on a été facilitateur, et c'était important, car c'était nouveau. Il y avait des réticences fortes, on est à l'interface et on peut rassurer des deux points de vue pour trouver une solution raisonnable, pour faire quelque chose qui soit suffisamment robuste, mais en leur faisant aussi part de la nécessité de gérer les contraintes* ».

Ces intermédiaires, positionnés à l'intersection des champs de l'économie appliquée et de l'action publique servent de traducteurs et de pacificateurs lors de ces réunions. Un membre de l'équipe de Martin Hirsch explique : « *Cet évaluateur c'est assez un Ayatollah, et ce n'était pas le plus souple, il n'y va pas avec le dos de la cuillère. Plusieurs fois, je lui ai dit « si tu ne veux pas qu'ils partent en claquant la porte, ne continue pas comme ça. Je comprends bien la réaction de la personne en face, et ce n'est pas la peine à un moment donné de nous flinguer le truc »*. Donc on a été beaucoup des « *go between* » [entremetteurs] pour faire les bons offices, et quelquefois ça n'a pas marché, ça a bien clashé ». Ils ont donc aidé à instaurer un consensus et participé au travail institutionnel cognitif pour convaincre les porteurs de projet du bien-fondé de ce protocole. Ces acteurs intermédiaires ainsi que les évaluateurs ont développé des discours et des outils pour « *faire avaler la chose* » – pour reprendre les termes d'un évaluateur – tout comme la permission de « *Jokers* » au tirage au sort⁸⁶³.

Ces situations conflictuelles lors de la phase de design ont principalement émergé des *relations cliniques* arrangées par des acteurs intermédiaires. Les relations cliniques constituées sur initiatives des évaluateurs ou des porteurs de projet et par des dynamiques de réseaux étaient beaucoup plus pacifiées au stade du design. Un des facteurs explicatifs est la sensibilité scientifique et l'appétence pour l'évaluation des porteurs s'étant inscrits dans cette démarche, ou la présence d'un acteur intermédiaire prenant à sa charge les contraintes et rendant « indolore » la randomisation. L'absence de conflit dans ces relations est aussi issue de la participation en amont du projet de l'évaluateur à la formulation du projet expérimental en tant que tel, ce qui, de fait, incorpore les exigences de protocole académique dès le départ⁸⁶⁴, comme notamment dans le cas des projets portés par les Universités, tels que « orientation en licence » avec l'université de Marne la Vallée, ou AQ3E avec l'université du Maine⁸⁶⁵.

Au fur et à mesure des expérimentations, les évaluateurs apprennent de cette première interaction, qu'elle soit collaborative et conflictuelle. Par la multiplication des interactions, ils choisissent davantage leur partenaire, identifient les projets porteurs de

⁸⁶³ Nous revenons sur ce procédé dans la partie 3.1.1.2 de ce même chapitre

⁸⁶⁴ Cette dimension est abordée en détail dans la partie 3 de ce chapitre portant sur la mise en œuvre.

⁸⁶⁵ Les évaluateurs de ces projets expliquent : « *C'est une expérimentation dont on a eu l'initiative. On a inventé le dispositif. L'idée c'était de tirer au sort des lycéens qui avaient exprimé le vœu d'être de rentrer à l'université de marne la vallée et de les inviter, (...) pour discuter avec un responsable de filière un enseignant chercheur de l'université pour que l'élève soit conforté ou au contraire soit ... ne soit pas conforté dans ses pré choix* ». (AQ3E), « *C'est une expérimentation qui a une histoire sympathique, au sens où c'est une idée entre nous, un soir, il nous faut un terrain, donc forcément universitaire, il y avait la logique de séparation avec l'expérimentateur, on fait partie d'un travail de recherche avec le CNRS Dans ce labo on a un co-auteur dans les testing et on leur proposé le projet* »

difficultés et s'appliquent à rendre le tirage aléatoire indolore. Les porteurs de projets apprennent aussi de cette interaction, mais ne sont pas en mesure d'appliquer ces apprentissages puisqu'ils ne peuvent qu'à titre exceptionnel porter plusieurs expérimentations consécutives.

2.2.2.2. Les apprentissages de la phase de cadrage, première contradiction d'intérêts d'apprentissages

L'étude des cas d'expérimentations, et particulièrement nos entretiens avec les différentes parties prenantes, se sont intéressés aux pratiques d'apprentissages lors de cette phase. Cette analyse révèle deux résultats. D'abord, que le modèle d'apprentissage scientifique de l'expérimentation aléatoire, lequel vise à un contrôle strict du design de l'expérimentation par les chercheurs, favorise les projets dont ces derniers ont l'initiative. Ensuite, lorsque ces projets n'étaient pas pensés par les chercheurs, on constate un dialogue difficile entre les objectifs de recherche des chercheurs, et les connaissances souhaitées par les porteurs de projets pour qui cette pratique de planification était nouvelle.

Lors de nos entretiens avec les chercheurs, ces derniers ont exprimé les enseignements qu'ils ont tirés de cette phase de design de l'expérimentation aléatoire. Ils témoignent d'une préférence pour les projets correspondant à leurs objectifs de recherche, leurs pratiques scientifiques et ne comportant pas d'enjeux centraux pour les acteurs de l'action publique. Pour le cas de la « Mallette des parents » par exemple, les scientifiques ont influencé la nature de l'intervention dès la conception du projet, un évaluateur explique : *« j'avais une revue de la littérature pluri-disciplinaire au moment où l'on interagissait avec les chefs d'établissement sur le thème de l'investissement des parents. (...) On a transféré cette information au responsable du rectorat qui s'occupait de concevoir le dispositif. »*

Le protocole scientifique a été d'autant mieux accepté par les acteurs de la mise en œuvre que les projets proposaient des interventions n'apportant pas un avantage important aux bénéficiaires. C'est dans cette logique que les expérimentations se sont portées sur des interventions « légères » comme l'accompagnement par exemple, le tutorat ou des réunions de parents d'élèves pour le cas de la « Mallette des parents », ce qui atténuait les réticences éthiques et/ou déontologiques vis-à-vis du groupe témoin. L'évaluateur de cette expérimentation explique : *« c'est aussi pour ça que la Mallette ça a bien marché, parce que ce n'était pas compliqué politiquement »*

Mais même si des interventions entrent dans cette catégorie politisée, le fait qu'elles soient portées par des acteurs de l'action publique qui sont convaincus du bien-fondé de cette démarche constitue un élément perturbateur pour le chercheur. Un évaluateur du TEPP résume : *« c'est eux [les porteurs] qui avaient eu l'idée, c'est eux qui croyaient dans leur dispositif et ça change tout dans la relation. Pour nous les scientifiques, ce qu'il y a de plus intéressant, c'est d'être à l'origine de l'idée de l'objet d'expérimentation »*. Les porteurs qui croient en leur dispositif sont beaucoup plus réfractaires à accepter d'en priver une partie des potentiels bénéficiaires et ils limitent aussi la validité externe des résultats puisque leur enthousiasme dans la mise en œuvre de l'intervention pourra être un facteur de l'impact estimé et potentiellement non reproductible en cas de généralisation.

Ces préférences entrent en contradiction avec les objectifs des porteurs de projets, opérateurs de l'action publique, dans la conduite d'expérimentations. Beaucoup estiment que ce travail sur le design a été un apport fondamental pour structurer le projet et ses

« process ». En effet un évaluateur du Crédoc explique : *« en tant qu'évaluateur, on exigeait des objectifs clairs, une population définie, donc ils se sont posés pas mal de questions »*. La plupart des porteurs estiment néanmoins que leur manque de culture évaluative ne leur a pas permis de réellement échanger avec l'évaluateur sur la pertinence du design scientifique vis-à-vis de leur projet. Le porteur de l'expérimentation de Tulle affirme : *« j'ai beaucoup appris. Moi j'en retire que du positif d'avoir travaillé avec des évaluateurs de ce niveau-là, c'est que du bonus. Après effectivement, au départ, on a appris au fur et à mesure. Maintenant, ça serait à refaire, on calibrerait beaucoup plus l'évaluation »*. Les porteurs de projets n'étaient pas habitués à passer autant de temps en amont sur la conception des outils de suivi et d'évaluation d'un projet. Un des membres du Cabinet de Martin Hirsch rapporte le constat commun des porteurs de projets lors d'une réunion thématique réunissant les acteurs des expérimentations du premier appel à projets du Haut Commissariat : *« la conclusion de beaucoup de porteurs de projets, c'est que le montage même de l'expérimentation leur avait demandé une énergie folle, qu'ils n'avaient pas prévu d'y passer autant de temps »*.

Ainsi, l'étude de l'amont de l'expérimentation nous renseigne à la fois sur les intérêts de ses parties prenantes et sur l'équilibre des forces au sein de cette relation clinique. Nos études de cas témoignent autant d'une diversité d'intérêts à conduire cette entreprise, que d'une diversité de profils d'acteurs. À l'issue de cette phase, un équilibre semble atteint dans la plupart des cas entre les intérêts de chacun. Mais si les chercheurs expérimentaux ont pour pratique traditionnelle de porter une attention soutenue à l'élaboration d'un protocole de recherche, estimant que ce dernier détermine en grande partie la qualité des résultats futurs, la pratique des acteurs de l'action publique est toute autre. De nombreux évaluateurs mobilisant des méthodes d'évaluation variées ont remarqué le manque de planification des projets d'action publique et particulièrement des projets expérimentaux. Les opérateurs de terrain et les managers suivent rarement un plan strictement établi, l'action publique s'élabore et se définit lors de son processus de mise en œuvre, et c'est ce dernier qui est jugé déterminant dans ce champ. C'est donc au stade de la mise en œuvre du projet que les acteurs de terrains interviennent et conduisent un travail institutionnel potentiellement déstabilisateur, mettant en danger l'intégrité du protocole scientifique élaboré en amont.

* * *

A l'issue de l'étude du début de la relation clinique de l'expérimentation aléatoire, plusieurs constats relatifs à notre cadre théorique ainsi qu'à nos hypothèses de recherche émergent. D'abord, l'étude des intérêts des parties prenantes de l'expérimentation aléatoire confirme que la décision de s'engager dans cette entreprise provient d'un ensemble d'enjeux liés d'une part, à des questionnements sur l'effectivité d'une intervention (*puzzling*) et d'autre part, à une recherche de consolidation ou d'amélioration de la position des parties prenantes dans leurs *champs d'action stratégiques* respectifs (*powering*). Néanmoins, il apparaît que même lorsque les motivations liées au *puzzling* sont en proportion importantes, par exemple pour les chercheurs, elles restent cependant motivées principalement par des enjeux liés au positionnement dans le champ d'action stratégique de ces derniers.

L'analyse de cette première phase de l'expérimentation aléatoire révèle aussi de nouveaux types d'interactions entre économistes appliqués et acteurs de l'action publique. Cette méthode organise pour la première fois l'interaction de ces deux groupes d'acteurs en amont de la mise en œuvre d'un dispositif. Lors de cette phase, les chercheurs tentent tant que possible d'influencer la conception du projet afin de le

rendre plus facilement évaluable. Face à eux, les porteurs de projets, dont l'accès au financement est sécurisé par leur appariement avec les évaluateurs pratiquant cette méthode souhaitée par le financeur, sont enclins à accepter plusieurs modifications de leur projet, notamment puisqu'ils n'en comprennent pas à ce stade systématiquement les implications concrètes. La mise en œuvre du projet va alors servir de révélateur de la solidité du compromis établi lors de la conception.

3. L'épreuve du terrain : la mise en œuvre de l'expérimentation aléatoire

« Une expérimentation randomisée, entre la très bonne idée sur le papier, d'une mesure ou même d'un protocole de randomisation très bien designé [anglicisme] et, 3 ans après, la récupération des données et des résultats, il s'est passé potentiellement beaucoup de choses. Il y a plein de vrais gens qui interagissent dans une expérimentation, que ce soit les traités ou que ce soit aussi les intermédiaires qui dispensent le traitement : les opérateurs, les gens qui travaillent concrètement à cela, il y a 1000 raisons qui font qu'à un moment donné la randomisation peut avoir sauté ou ne pas avoir été appliquée, que les règles n'aient pas été parfaitement appliquées et auxquels cas les données que l'on récupère ne valent rien ». Comme le suggère ces propos d'un évaluateur du Crest, la mise en œuvre de l'expérimentation constitue la phase la plus sujette au travail institutionnel des parties prenantes de l'expérimentation. Cette période, allant de deux à trois ans, nécessite une interaction permanente entre les porteurs de projets et les équipes d'évaluation qui doivent, au-delà de leurs ambitions de départ, faire face aux réalités de terrain. Alors que cette étape fait appel à des pratiques connues par la plupart des porteurs de projets, elle se présente comme une découverte pour les chercheurs/évaluateurs qu'aucune autre méthode d'évaluation n'a amenés à être aussi proches de l'action ainsi qu'à la suivre dans son développement. De plus, la structure du protocole nécessaire à l'apprentissage scientifique expérimental s'avère particulièrement contraignante pour les opérateurs de terrain dont les pratiques menacent l'intégrité.

Tout apprentissage suggère un processus d'apprentissage, un cheminement, l'emprunt d'une voie, d'une méthode amenant d'un état de connaissance ou de perception à un autre. Ce chemin est variable selon les pratiques de champs. Le champ scientifique se distingue par un processus d'apprentissage dont la conformité à une méthode rigoureuse de construction de connaissance est primordiale. À l'inverse, les acteurs du champ de l'action publique ont une obligation d'action les entraînant à apprendre par et pendant la mise en œuvre, et à ajuster cette dernière au vu de leur expérience. La méthode expérimentale par assignation aléatoire, en tant qu'institution scientifique d'apprentissage, requiert une conformité forte de la mise en œuvre au protocole préalablement défini, incluant un contrôle de l'administration du « traitement », ainsi que de la composition des groupes de comparaison. Ces exigences ont été sujettes au plus fort travail institutionnel, reflétant ainsi des pratiques en opposition radicale. Deux enjeux centraux émergent de l'analyse de nos cas d'étude. Le premier est l'organisation et le maintien du tirage au sort des bénéficiaires pour la constitution des groupes de comparaison, ce qui constitue la clé de voûte de cette méthodologie. Le deuxième enjeu, qui semble secondaire, nous apparaît néanmoins comme l'élément le plus problématique de cette entreprise. Il réside dans le contrôle et la connaissance par les chercheurs du traitement *réellement effectué*, de l'intervention expérimentale.

3.1. Confrontation des types de travail institutionnel autour de l'assignation aléatoire

Le respect de l'allocation aléatoire des individus au sein du groupe test et du groupe témoin constitue la clé de voûte garantissant la rigueur des résultats scientifiques produits, elle est donc essentielle pour l'évaluateur/chercheur. Cependant, cette condition est justement celle qui interfère le plus avec la pratique des acteurs de la mise en œuvre. Elle a été l'objet d'un travail de contrôle conséquent par les évaluateurs afin de maintenir ce pré-requis de l'expérimentation aléatoire. Ce travail institutionnel permanent de l'évaluateur, qu'il consiste en une ingénierie pratique imposant ce pré-requis, ou dans un travail cognitif d'explication de son importance, a abouti à de nouveaux compromis avec le porteur de projets. En effet, cette entreprise a rencontré un travail institutionnel de déstabilisation des acteurs de terrains qui, soit par un travail institutionnel passif de conservation de leurs pratiques habituelles, soit par un travail actif de détournement, ont régulièrement perturbé ce protocole lorsque ce dernier n'avait pas été pleinement accepté.

3.1.1. Le travail institutionnel de maintien de l'assignation aléatoire

Le travail de contrôle du tirage aléatoire des bénéficiaires de l'intervention évaluée a constitué une source de travail conséquente pour l'évaluateur tout au long de la mise en œuvre du projet. Elle est plus ou moins facile à contrôler selon le niveau où survient le tirage aléatoire et selon le nombre d'intermédiaires de terrain pouvant interférer avec cette pratique. Au fur et à mesure des expérimentations, les évaluateurs ont tenté d'organiser ce tirage au sort en amont de la mise en œuvre afin de réduire cette charge de travail. Néanmoins, pour les premiers cas d'utilisation de la méthode aléatoire, où chercheurs et opérateurs étaient novices dans l'organisation de cette répartition, on retrouve pour de nombreux cas des aménagements à cet impératif méthodologique.

3.1.1.1. Le travail de renforcement

« Il y a quelque chose de particulier à cette approche, c'est qu'on doit aider à faire bouger des forces politiques et sociales. C'est souvent extrêmement compliqué, lourd et prenant, et c'est particulier à ce champ de recherche. En sciences sociales, il n'y a pas de difficultés équivalentes à celles auxquelles on doit faire face dans la confrontation opérationnelle avec la réalité sociale et politique. C'est difficile et fatigant ». Ces propos d'un évaluateur de l'École d'Économie de Paris témoignent du lourd travail institutionnel nécessaire de la part des chercheurs afin de garantir l'intégrité du protocole expérimental. Ce travail est double. Il comporte d'une part un travail d'ingénierie et de contrôle sur le terrain, et d'autre part, un travail cognitif de persuasion des acteurs de la mise en œuvre de la nécessité de se plier à ce protocole.

La recherche expérimentale s'appuie sur une manipulation de l'objet par le chercheur qui crée une situation de comparaison artificielle permettant de tester une théorie. Son application à des sujets humains, hors laboratoire, requiert un travail conséquent de contrôle tout au long de l'expérimentation. Pour les chercheurs qui se sont investis dans cette entreprise pour nourrir leurs travaux de recherche, cette conformité au protocole est essentielle, comme en témoigne un chercheur du TEPP : *« s'ils n'ont pas appliqué la randomisation comme on le souhaitait, c'est foutu, et on ne peut rien en faire. Et là, on a perdu 3 ans de recherche,(...) trois ans de labeur terrible ! ».* Ce travail consiste

d'abord en une vérification systématique, par l'évaluateur, que les bénéficiaires sélectionnés par le tirage au sort soient bien ceux qui avaient reçu la prestation. La proximité géographique des chercheurs avec la mise en œuvre de l'expérimentation ainsi que sa concentration sur un seul site facilitent grandement ce travail⁸⁶⁶. L'ampleur de ce travail nécessite une taille d'équipe de recherche importante. Ce travail est principalement effectué par les chercheurs que nous avons identifiés comme *satellites* qui peuvent être épaulés par des assistants de recherche. Cette gestion de ressources humaines est une activité nouvelle pour les équipes de chercheurs. Pour les équipes de l'École d'Économie de Paris, elle est effectuée par le J-PAL qui prend à sa charge le recrutement et la formation de ces assistants. Un des porteurs du projet de la « Mallette des parents » explique le travail de contrôle de ces assistants : *« ils [l'équipe d'évaluation] avaient une armée d'assistants de recherche pour aller assister à tous les débats et vérifier que les parents dans les débats étaient ceux tirés au sort »*. Ce travail est beaucoup plus lourd pour les équipes du TEPP, du Crédoc et du Céreq ne bénéficiant pas de cet appui. Une des solutions alternatives pour ces chercheurs est alors, comme l'explique un évaluateur du Crédoc, de cadrer le plus possible l'action sur le terrain : *« vu qu'on avait des procédures très formalisées, on a laissé très peu de marge de manœuvre au porteur de projets pour tricher par rapport à l'aléatoire »*.

L'ampleur de ce travail peut parfois mener à un mélange des rôles entre porteur de projets et évaluateurs. L'évaluateur est amené à intervenir régulièrement dans la mise en œuvre afin de garantir la production de résultats, et donc à appuyer le travail des porteurs de projets. Les équipes d'évaluation ont du notamment s'assurer que le projet concernait assez de bénéficiaires afin d'obtenir une puissance statistique suffisante pour en mesurer l'impact. Cette interférence était d'autant plus présente dans les cas où l'évaluateur avait participé à la conception du projet. L'évaluateur du projet AQ3E explique cette confusion des rôles : *« Quand c'était difficile, c'était plus ma casquette d'enseignant à la fac qui m'a fait tenir. Je croyais en l'utilité sociale de ce projet et j'avais plus une casquette d'expérimentateur que d'évaluateur. Je me disais, il faut laisser sa chance à ce dispositif, car si ça marche, c'est très important pour les étudiants »*. D'autre part, les porteurs de projets, eux-mêmes convaincus de la nécessité de ce protocole pour établir l'efficacité de leur projet, ont régulièrement appuyé le travail institutionnel des évaluateurs de maintien du protocole évaluatif. Par exemple, dans le cas de l'expérimentation de la Mission Locale de Tulle, la chargée de mission de la Mission Locale a servi d'intermédiaire entre les évaluateurs et les acteurs de terrain afin d'assurer le respect du tirage au sort⁸⁶⁷. Ca a été aussi le cas pour les expérimentations menées par le rectorat de Créteil ou le projet « APIPAQ »⁸⁶⁸.

Le poids de ce travail de contrôle du respect de l'allocation aléatoire a été source d'apprentissage pour les chercheurs qui, au fil des expérimentations qu'ils conduisaient, ont cherché à maîtriser le respect du tirage au sort en amont. Un évaluateur des TEPP explique cette démarche : *« C'est très très lourd, on réfléchit de plus en plus à des protocoles qui allègent notre implication chez l'expérimentateur. À titre personnel, je ne veux pas revivre ça. On a identifié que c'est l'histoire des intermédiaires. Quand tout*

⁸⁶⁶ Un évaluateur de la mallette des parents explique : *« c'est plus facile d'observer, de discuter avec les acteurs quand ils sont proches et de les voir en face à face, donc on essaie de se déplacer beaucoup sur le terrain de les appeler beaucoup au téléphone pour répondre à leur question et je pense que c'est la clé de la réussite, »*

⁸⁶⁷ Un des porteurs de ce projet résume : *« c'est moi qui me suis arrangé pour que ça marche »*

⁸⁶⁸ L'évaluateur du projet Apipaq explique : *« Il fallait convaincre les expérimentateurs du bien-fondé de la méthode pour qu'après ils fassent le nécessaire pour qu'après de leurs équipes ce soit bien accepté », ou on lit dans le rapport d'évaluation de l'internat de Sourdon un remerciement pour le « soutien indéfectible du rectorat de Créteil et de la DGESCO »* ECOLE D'ECONOMIE DE PARIS, *Les effets de l'internat d'excellence de Sourdon sur les élèves bénéficiaires : résultats d'une expérience contrôlée*, Rapport d'évaluation, Fonds d'expérimentation pour la jeunesse, 2013, p. 21.

dépend de la qualité des gens, de leur implication, alors qu'ils ne sont pas porteurs de projets, c'est très compliqué. ». D'abord, les risques peuvent être limités en modifiant le niveau de la répartition aléatoire. En comparant au niveau des classes au lieu des individus, comme cela a été le cas pour la « Mallette des parents », ou en faisant intervenir le tirage au sort en amont de l'intervention comme dans le cas de la mission locale de Tulle. Le tirage aléatoire devenait alors moins visible, voire invisible, et ne venait plus perturber les pratiques d'interactions entre acteurs de la mise en œuvre de l'action publique et bénéficiaires. Ce positionnement du tirage aléatoire permet à la fois de ne pas perturber le travail des opérateurs de terrain et dans certains cas, une mise en œuvre de l'expérience « en aveugle ». Par exemple, dans le cas de l'expérimentation de Tulle, les jeunes du groupe test étaient orientés lors de la phase 1 vers les accompagnements au contrat d'apprentissage, mais les membres du groupe témoin pouvaient aussi, à leur initiative, bénéficier de cet accompagnement. Pour reprendre les termes du porteur : *« La sélection aléatoire a marché parce que ça n'a pas généré d'accompagnement différent ça a généré le rajout d'une étape à un moment donné qui a été fait par des conseillers, avec qui j'ai bien discuté, qui intervenaient au CFA, mais pour les autres ça ne changeait rien (...), ils n'avaient pas conscience de la moulinette [système informatique de tirage aléatoire], les conseillers ne savaient pas qui étaient dans l'expérimentation ou pas. »*

Outre ce travail institutionnel de contrôle matériel, les chercheurs doivent aussi tout au long de l'expérimentation, mais particulièrement au début de sa mise en œuvre, procéder à un travail de nature cognitive. Nous avons déjà évoqué précédemment l'ampleur du travail cognitif conduit par les évaluateurs ainsi que les financeurs des expérimentations afin de promouvoir la supériorité de la méthode par assignation aléatoire pour établir quantitativement l'impact d'un projet. Néanmoins, si les porteurs de projets peuvent à ce stade être convaincus, il reste à persuader les opérateurs de terrain qui sont au contact des bénéficiaires et concrètement chargés de mettre en œuvre le dispositif. Plus ces acteurs sont convaincus de l'importance du respect du protocole, moins le travail de contrôle des évaluateurs sera important. Un évaluateur du Crest explique : *« L'enseignement principal, c'est qu'il faut que tout le monde soit d'accord, il faut le faire en particulier avec les gens qui concrètement font le travail. S'ils ne sont pas d'accord pour respecter le protocole expérimental, on ne peut pas les obliger, on ne peut pas les menacer de les licencier avec un fusil sur la tempe. Ce n'est pas envisageable. S'ils décident de mettre en l'air le protocole, ils ont les moyens de le faire. Donc c'est vraiment indispensable qu'ils soient convaincus que c'est ça [le tirage aléatoire] qu'il faut faire »*.

Ce travail de « pédagogie » s'effectue au cours de réunions parfois conflictuelles où l'équipe d'évaluation rappelle la nécessité de l'emploi de cette méthode pour produire des résultats crédibles et légitime, à savoir le recours au tirage au sort. Un des argumentaires résidait sur la réputation internationale de rigueur de cette méthode ; un porteur de projets rappelle ce discours : *« On leur disait « là on vous propose de participer à un protocole d'évaluation qui suit des standards internationaux, jamais dans votre vie professionnelle vous rencontrerez à nouveau ce protocole » »*. Un argumentaire commun est préparé alors autour de la pertinence du tirage aléatoire pour contrer les oppositions de nature éthique. Cet argumentaire explique d'une part, que les inégalités de traitement existent déjà entre les publics, comme dans les voyages scolaires par exemple qui sont conduits uniquement pour certaines classes.

D'autre part, il rétorque que, du fait que l'efficacité du traitement ne soit pas encore

établie, on ne prive pas forcément les non-bénéficiaires d'une action bénéfique⁸⁶⁹. Enfin, il explique que le travail de sélection des accès à l'aide, conduit par les conseillers, n'est pas forcément moins discriminant qu'un tirage aléatoire. C'est notamment l'argumentaire qui a été développé dans le cadre de l'expérimentation des permis de conduire, comme l'explique un des évaluateurs : « *On ne pouvait pas donner de l'argent à tout le monde, et la rationalisation la plus juste c'est peut-être le tirage au sort* ». Un évaluateur du Céreq renchérit : « *le conseiller pense faire au mieux selon ce qu'il pense de la personne, mais il n'est pas omniscient. Il fait une sélection sur la base de critères parfois subjectifs, et ce n'est pas forcément légitime.* »

Néanmoins, malgré ces différents travaux de maintien de l'institution, les évaluateurs apprennent à concéder à plusieurs compromis afin de rendre acceptable le tirage aléatoire et que le protocole ne soit pas complètement mis de côté.

3.1.1.2. Les compromis des évaluateurs

Plusieurs compromis ont été mis en place lorsque les résistances des acteurs de terrain menaçaient trop fortement l'intégrité de la procédure. Elles ont servi de soupapes pour que s'exprime la subjectivité des acteurs de la mise en œuvre.

Un des compromis a consisté à accepter que les jeunes en situation les plus précaires puissent accéder au dispositif par le biais de « joker ». Par exemple, dans l'expérimentation AgiLoJe, chaque porteur bénéficiait de 5 jokers, qui faisaient ensuite sortir le bénéficiaire de l'évaluation. L'évaluateur des permis de conduire explique : « *On a mis en place une chose qui les a beaucoup rassurés, c'était la logique du joker ! Si pour un bénéficiaire c'était indispensable et que sa vie était foutue, il pouvait donner la subvention sans tirage au sort, mais il sortait de l'expérimentation et de l'évaluation. Ça les a rassurés, ils restaient actifs et maîtres s'il y avait un cas de force majeure, mais en fait cela a été très peu utilisé.* » Dans le cas de l'expérimentation de la mission locale de Beaune, tout jeune du groupe témoin connaissant une rupture de contrat a été immédiatement pris dans le dispositif et a bénéficié d'un accompagnement adapté, ce qui a été le cas pour 55 jeunes⁸⁷⁰.

Aussi, dans plusieurs situations, des évaluateurs ont accepté, à titre exceptionnel, que des non-bénéficiaires, s'ils en exprimaient l'envie, puissent bénéficier du dispositif et sortent de l'expérimentation comme par exemple dans l'expérimentation sur l'accompagnement des RMistes où 6% du groupe témoin ont fait cette requête⁸⁷¹. Des dérogations ont aussi été allouées dans le cas où des bénéficiaires et non bénéficiaires étaient en contact, par exemple via un apprentissage dans une même entreprise, comme dans le cas de l'expérimentation de Beaune.

Un autre compromis a consisté à limiter le tirage aléatoire à une première période du projet, comme cela a été le cas pour l'accompagnement des RMistes qui a limité le tirage aléatoire aux 18 premiers mois. Enfin un autre compromis rencontré dans cette

⁸⁶⁹ L'évaluateur du projet Apipaq explique : « *les personnes avec qui on a discuté au départ ce n'était pas les conseillers de Mission Locale qui allaient directement être en prise avec les jeunes, ce n'était pas les personnes à l'éducation nationale qui gèrent les décrocheurs directement, (...) ça n'allait pas impacter leurs pratiques quotidiennes alors que quand on passe aux opérateurs là ça devient plus compliqué de dire un tel tu peux lui proposer d'aller sur le dispositif et l'autre non et là du point de vu humain ça a été un peu plus dur, parce qu'il y a une autre chose que les gens n'arrivent pas à comprendre, c'est qu'ils partent du principe que le dispositif a forcément un impact positif et donc comme il a un impact positif donc priver des individus de ce dispositif n'est pas éthique* ».

⁸⁷⁰ IREDU/CNRS, *Inscrire les contrats en alternance dans une logique de parcours sécurisé*, Rapport d'évaluation, Ministère de l'Éducation nationale, de la Jeunesse et de la Vie associative, 2012, p. 42.

⁸⁷¹ Laurent LEQUIEN, *Essais d'évaluation de politique publique dans les champs de l'éducation, de la santé et des politiques d'emploi*, Paris, Ecole des Hautes Etudes en Sciences Sociales, 2011, p. 25.

même expérimentation prévoit non pas un tirage au sort à parts égales entre le groupe témoin et le groupe test, mais une proportion de $\frac{1}{4}$ de témoin et $\frac{3}{4}$ test.

Bien sûr, ces aménagements compliquent par la suite l'analyse effectuée par les évaluateurs, aussi parce que toutes ces exceptions sont autant d'individus sortis de l'analyse de données, réduisant la puissance statistique de l'expérimentation.

Néanmoins, dans certains cas, le travail institutionnel des évaluateurs ou leurs compromis pour convaincre les acteurs de la mise en œuvre n'a pas été efficace et les évaluateurs ont été confrontés à un travail institutionnel de déstabilisation, actif et volontaire, de la part des opérateurs. Au-delà de ce travail institutionnel de déstabilisation par les acteurs de la mise en œuvre, les évaluateurs ont aussi parfois fait face à d'autres éléments perturbateurs, de type organisationnel, directement liés à l'expérimentation en elle-même, et susceptibles de compromettre leur protocole et donc la production de résultats probants.

3.1.2. Les facteurs de déstabilisation du protocole expérimental d'assignation aléatoire

Plusieurs éléments concourent à la déstabilisation du protocole expérimental et menacent la production de résultats scientifiques. D'abord, un ensemble de facteurs n'est pas issu d'un travail institutionnel de déstabilisation, mais tient à la nature expérimentale du projet, comme notamment un important non-recours des individus du groupe test au traitement, ou à l'inverse un recours au traitement des individus du groupe témoin dans le cadre d'un projet déjà existant. Par ailleurs, nous étudions pourquoi la pratique du tirage aléatoire a pu s'inscrire en porte à faux avec les pratiques du champ de l'action publique et notamment des acteurs de l'action sociale, et nous présenterons les types de travail institutionnel de déstabilisation dans lesquels ces derniers se sont engagés.

3.1.2.1. Les éléments de déstabilisation du protocole propres à l'expérimentation

Un problème majeur rencontré par les évaluateurs est qu'un faible nombre de bénéficiaires participent à l'expérimentation. Cet élément perturbateur peut être rencontré par tout projet innovant, même si ce dernier n'est pas évalué par la méthode aléatoire, mais il est particulièrement présent avec ce protocole puisqu'il requiert que la moitié des bénéficiaires potentiels n'y aient pas accès afin de constituer le groupe témoin. De plus, un nombre suffisant de bénéficiaires est nécessaire à cette méthode afin d'atteindre un coefficient de puissance statistique permettant d'estimer l'impact sans aléa statistique important. Une moitié de l'échantillon devant être témoin, cela nécessite d'identifier le double de bénéficiaires potentiels vis-à-vis du nombre d'individus pouvant être traités⁸⁷². Cette condition nécessite d'avoir une très bonne connaissance du public ciblé, de savoir comment le mobiliser et d'avoir une capacité à délivrer le traitement à un nombre important de bénéficiaires, conditions rarement réunies dans le lancement de nouvelles interventions⁸⁷³. Ce problème est d'autant plus fort pour

⁸⁷² Pour exemple, afin d'obtenir un groupe test constitué de 100 bénéficiaires effectifs de l'intervention, il est nécessaire au préalable d'identifier 200 bénéficiaires potentiels qui seront ensuite répartis aléatoirement en deux groupes de 100, un test, et un témoin.

⁸⁷³ Cette problématique est résumée par les évaluateurs du projet apipaq : « la population cible incluse en pratique par l'expérimentation est sensiblement différente de la population théorique, de nombreuses restrictions ayant été effectuées » Dominique MAILLARD, Virginie MORA, Pascale ROUAUD, Michel THERY, Patrick VENEAU et Tania LEHBERGER, « Évaluation d'un dispositif d'accompagnement de jeunes en rupture scolaire », *Net.Doc* 2012.97, 14.

l'expérimentation aléatoire puisque selon la nature du tirage aléatoire, il n'est parfois pas possible d'organiser une communication publique autour du projet pour en augmenter le recours⁸⁷⁴.

Cette difficulté opère à deux niveaux, celui de l'identification de la population de base et ensuite celui du groupe test. De nombreuses expérimentations s'adressent à des publics fragiles difficiles à identifier, comme les décrocheurs en milieu scolaire par exemple, population échappant à un recensement systématique. Une tâche préliminaire au tirage aléatoire consistait donc à identifier ces publics, les contacter et qu'ils acceptent de se porter candidat et de se soumettre au tirage aléatoire. Lorsque des expérimentations portent en partie sur le mode de sollicitation des publics, il est impossible de conduire une campagne de communication qui brouillerait alors les estimations de l'impact différencié des modes de sollicitations entre le groupe témoin et le groupe test. Cette promotion est aussi impossible quand les évaluateurs veulent contrôler précisément le profil des potentiels bénéficiaires. Cette situation, poussée à l'extrême, a pu mener à l'abandon de l'expérimentation du projet, comme ce fut le cas pour une expérimentation de lutte contre le décrochage universitaire portée par l'université de Marne la Vallée et évaluée par le Céreq. L'évaluateur explique : *« ils voulaient diffuser de façon assez large sur leur site, placarder dans l'université, informer le maximum d'étudiants sur la démarche d'expérimentation. Nous on a dit « non, il nous faut une population témoin, il faut que l'on choisisse les gens selon certains critères, il faut faire une pré-enquête pour trouver des vrais sortants ».(...) On avait mis des conditions telles qu'on avait touché très peu de public et ce public s'évaporait dans la nature et personne n'a suivi leur stage. Ca été une catastrophe pour eux »*. Dans la précipitation de la constitution de la relation clinique entre évaluateur et porteur de projets, ces derniers ont pu envisager que les effectifs de la population de bénéficiaires potentiels seraient suffisants et peuvent se rendre compte finalement, à l'épreuve du terrain, que l'atteinte du nombre d'individus requis est plus problématique⁸⁷⁵. Cette difficulté a été rencontrée dans de nombreux cas d'expérimentations comme celui de Tulle, où l'expérimentation a été étendue à d'autres territoires afin d'atteindre le nombre suffisant de bénéficiaires potentiels. Une autre solution a été d'élargir la période d'expérimentation, comme dans le projet « Rebond »⁸⁷⁶.

De plus, à la difficulté d'identifier une population de potentiels bénéficiaires suffisante, s'ajoute une autre difficulté si les individus du groupe test ont le choix d'avoir recours ou non au traitement. Ce faible recours ou *take-up*, pour reprendre le terme consacré en économie, peut avoir lieu sur l'ensemble du dispositif, comme pour l'expérimentation sur l'accompagnement renforcé où le recours n'a été que de 15% pour le groupe test ; ou pour le projet Apipaq où seulement 29% des jeunes orientés vers le dispositif y ont eu

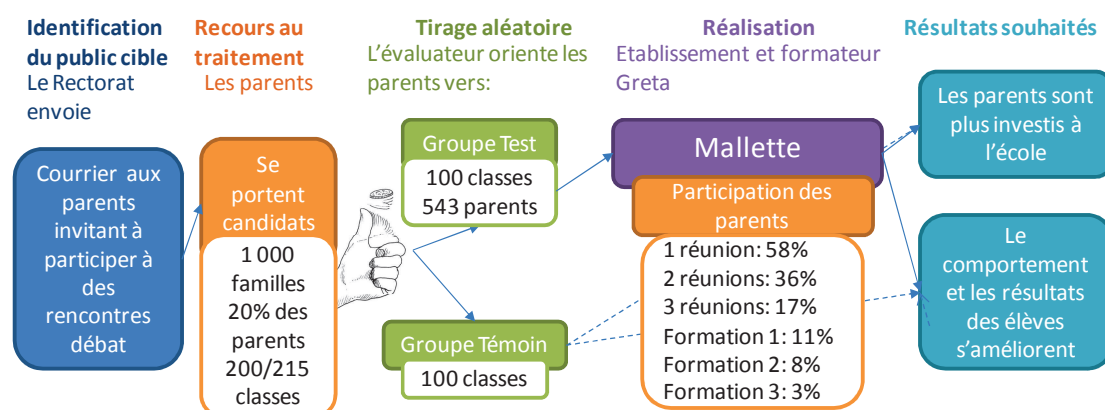
⁸⁷⁴ Pour quelques expérimentations comme celle sur les permis de conduire ou les internats d'excellence il a été possible de communiquer autour du projet puisque les candidats se présentaient et ensuite étaient tirés au sort.

⁸⁷⁵ Un évaluateur du Crédoc revient sur les raisons de l'abandon du protocole aléatoire pour une expérimentation menée avec la fondation des Apprentis d'Auteuil : *« Lors du projet, les porteurs ont intérêt à dire « oui, oui » car ils ont besoin d'un évaluateur. Une fois qu'il est sélectionné, on se rend compte que certaines des conditions indispensables sur le papier ne l'étaient pas sur le terrain. Lors de la phase de négociation, certaines choses ont été avancées qui lors de la phase de mise en œuvre n'ont pas pu être maintenues (...) On a travaillé sur l'aléatoire avec eux. Ils s'attendaient à ce que les réunions d'information soient beaucoup suivies, mais peu de jeunes ont participé, et peu voulaient rentrer dans le projet. Au bout de 6 mois on s'est rendu compte qu'on était en train de tuer le projet et d'en prendre un pour deux c'était très décourageant pour les personnes qui accompagnaient les jeunes et qui déjà faisaient beaucoup d'effort pour les faire entrer dans le dispositif »*

⁸⁷⁶ On lit dans le rapport d'évaluation : *« Afin de nous rapprocher des objectifs quantitatifs fixés au début de l'expérimentation, nous avons décidé, en accord avec la structure expérimentatrice, d'étendre la période d'inscription dans le dispositif jusqu'au mois de décembre 2010 »* CREST, Note d'étape sur l'évaluation, REBOND, pour les étudiants décrocheurs de l'enseignement supérieur, Mission d'animation du Fonds d'expérimentations pour la jeunesse, 31 décembre 2010.

recours⁸⁷⁷. Si le dispositif consiste en plusieurs interventions, les sorties du dispositif peuvent survenir aussi au cours du déploiement du programme, comme cela a été le cas pour l'expérimentation de la « mallette des parents ». Le diagramme ci-dessous représente ce dispositif et la diminution de la participation des parents au cours d'année

Figure 14: Représentation graphique de l'ingénierie de l'expérimentation "Mallette des parents" et du nombre de bénéficiaires



Quand s'ajoute une difficulté d'identification d'une population suffisante, et un faible recours au dispositif, l'intégrité scientifique des résultats peut être fortement menacée. Ce type de situation est d'autant plus compliqué dans des expérimentations où les porteurs de projets sous-traitent la mise en œuvre. Les opérateurs sous-traitants sont unis aux porteurs par une contractualisation établissant un nombre de bénéficiaires et ces derniers ont donc recruté du personnel en fonction de cette estimation du nombre de bénéficiaires. Cette difficulté a été rencontrée dans le cas de l'accompagnement des RMIstes et de l'expérimentation « jeunes diplômés »⁸⁷⁸. L'évaluateur de cette dernière expérimentation explique : « on avait calibré [le projet] sur les flux des années précédentes, mais les flux étaient moins importants. Moins de jeunes au chômage. On a dit « il faudra viser 50% de take up parce que sinon on ne mesure rien ». Donc on a dit « on va donner aux OPP [opérateurs privés de placement] uniquement 2 fois plus de jeunes que nécessaire » parce qu'eux, les OPP, ils avaient 2 000 jeunes à accompagner sur un an, donc ils s'étaient outillés pour ça, ils avaient recruté des gens. On voulait ne pas trop écrémer pour avoir une population suffisamment représentative des jeunes et puis surtout qu'on ne perde pas de puissance statistique. Ce qu'on a fait, finalement, c'est qu'on a ouvert d'autres zones où ce n'était pas soumis à l'expérimentation, où ils étaient tranquilles, ils pouvaient gagner de l'argent là, et ils n'étaient plus pris à la gorge à cause de nous ».

Enfin, la méthode aléatoire est particulièrement difficile à appliquer dans un contexte où un projet existant déjà, préalablement à l'expérimentation formelle, fait appel à des pratiques similaires à celui du dispositif expérimenté. Ces activités hors du cadre juridique expérimental sont alors englobées dans le droit commun et ne peuvent de fait, être refusées à des individus souhaitant en bénéficier. Cet élément peut amener de

⁸⁷⁷ On lit dans le rapport : « Parmi les jeunes du groupe « intervention », 149 ont rejoint un EAP soit 29%. Les formations suivies par les jeunes depuis leur date de décrochage scolaire ne diffèrent pas significativement entre les deux groupes « intervention » et « témoin » Dominique MAILLARD et al., « Evaluation d'un dispositif d'accompagnement de jeunes en rupture scolaire », *Net.Doc* 2012.97, 4.

⁸⁷⁸ Un document de synthèse de cette expérimentation explique « Selon les fichiers transmis par les OPP, ce sont 9 890 jeunes sur 31 854 orientés vers les OPP entre septembre 2007 et septembre 2008 qui ont finalement signé un contrat d'accompagnement, soit un taux d'adhésion de 31 %. Bruno CREPON, Esther DUFLO, Marc GURGAND, Roland RATHELOT et Philippe ZAMORA, « L'accompagnement des jeunes diplômés demandeurs d'emploi par des opérateurs privés de placement: les enseignements d'une évaluation », *Dares Analyses* 2011.94.

nouveaux aménagements au sein du tirage au sort et diminuer la taille de l'échantillon d'individus répartis aléatoirement entre les deux groupes. Par exemple dans l'évaluation menée par la Mission Locale de Beaune, l'existence préalable du projet sur une des trois Missions Locales participantes a amené les évaluateurs à accepter qu'aucun des potentiels bénéficiaires du dispositif identifiés par la Mission Locale de Beaune ne soit orienté vers le groupe témoin, ce qui a contribué à exclure 131 jeunes de l'échantillon des évaluateurs⁸⁷⁹. Cette même expérimentation a aussi pâti de procédures déjà existantes, mais non efficaces pour la composition de la population mère. Cette dernière était constituée sur la base du numéro d'enregistrement de Cerfa, actant l'enregistrement du contrat d'apprentissage, mais peu à peu les expérimentateurs se sont aperçus que l'enregistrement officiel pouvait arriver en moyenne trois mois après l'élaboration du contrat⁸⁸⁰. C'était autant de temps perdu de suivi par la Mission Locale et faisait perdre de l'efficacité au dispositif. Les expérimentateurs ont donc finalement arbitré pour constituer la population mère directement à partir des élèves recensés au CFA⁸⁸¹. Ces difficultés combinées ont abouti à une ingénierie complexe du tirage au sort comme en témoigne ce tableau récapitulatif issu du rapport d'évaluation⁸⁸².

Figure 15: Ingénierie du tirage aléatoire pour l'expérimentation "inscrire les contrats en alternance dans une logique de parcours sécurisé », porté par la Mission Locale de Beaune

Sélection des deux groupes	Mode de sélection	- tirage aléatoire (randomisation)
	Unité de sélection	- année 1 : numéro de Cerfa de l'apprenti-e pour le secteur HCR sauf pour la Mission locale de Beaune où tous les apprentis bénéficient du tutorat renforcé - année 2 : jour de naissance de l'apprenti-e pour les secteurs HCR et commerce sauf pour la Mission locale de Beaune où tous les apprentis du secteur HCR bénéficient du tutorat renforcé ; localisation de l'entreprise pour le secteur de la viticulture pour les Missions locales de Beaune et de Dijon
	Critères de sélection	- année 1 : • Mission locale de Beaune : tutorat renforcé pour tous les apprentis du secteur HCR • Missions locales de Dijon et de Montbard : numéro de Cerfa pair → tutorat renforcé (bénéficiaire), numéro de Cerfa impair → information (témoin) pour le secteur HCR - année 2 : • Mission locale de Beaune : tutorat renforcé pour tous les apprentis du secteur HCR ; jour de naissance pair → tutorat renforcé, jour de naissance impair → information pour le commerce • Missions locales de Dijon et de Montbard : jour de naissance pair → tutorat renforcé, jour de naissance impair → information pour les secteurs HCR et commerce • Missions locales de Beaune et de Dijon : entreprise localisée en Côte-d'Or → tutorat renforcé, entreprise localisée hors Côte-d'Or → information pour le secteur de la viticulture

⁸⁷⁹ IREDU/CNRS, *Inscrire les contrats en alternance dans une logique de parcours sécurisé*, Rapport d'évaluation, Ministère de l'Education nationale, de la Jeunesse et de la Vie associative, 2012, p. 49.

⁸⁸⁰ Cette pratique devrait être maintenant moins courante depuis que la loi Cherpion impose un enregistrement des contrats sous 15 jours et préconise le développement d'un enregistrement numérique des contrats. *LOI n° 2011-893 du 28 juillet 2011 pour le développement de l'alternance et la sécurisation des parcours professionnels*, (2011) 2011-893.

⁸⁸¹ MISSION LOCALE DE BEAUNE, *Note d'étape sur l'expérimentation, Inscrire les contrats en alternance dans une logique de parcours sécurisé*, Mission d'animation du Fonds d'expérimentations pour la jeunesse, 31 décembre 2010.

⁸⁸² IREDU/CNRS, *Inscrire les contrats en alternance dans une logique de parcours sécurisé*, Rapport d'évaluation, Ministère de l'Education nationale, de la Jeunesse et de la Vie associative, 2012, p. 29.

Ainsi, sans l'intervention d'aucun acteur, des contextes expérimentaux peuvent déjà constituer des risques pour la production de résultats scientifiques, notamment par la difficulté à atteindre des échantillons suffisants pour garantir une puissance statistique établissant la robustesse des résultats de l'évaluation. De plus, dans de nombreux cas d'expérimentation, les acteurs de terrain se sont engagés dans une variété de types de travail institutionnel déstabilisateur de l'intégrité de cette institution.

3.1.2.2. Le travail institutionnel volontaire de déstabilisation

Avant d'aborder le travail institutionnel de déstabilisation conduit par les acteurs de terrain de la mise en œuvre, il est nécessaire de revenir sur les raisons de ces positionnements. Cette opposition ne tient pas uniquement à une préoccupation éthique, mais surtout à la remise en cause, par le tirage aléatoire, de l'identité professionnelle ainsi que du savoir-faire des acteurs de la mise en œuvre. Ces ressentiments ont pris une intensité variable selon les identités de ces derniers et leur portage du projet évalué. Ils ont abouti à des actions de déstabilisation lorsqu'ils intervenaient comme intermédiaire et média entre le dispositif et le bénéficiaire. Un évaluateur du Crest revient sur cette pratique : *« ils [les porteurs de projets] s'engagent dans une opération, ils ne se rendent absolument pas compte des enjeux scientifiques, pas de l'enjeu, mais des exigences scientifiques. Il faut vraiment que l'évaluation soit complétement circonscrite, qu'elle soit complétement nouvelle et que seuls les jeunes du groupe test en bénéficient. »*

Au-delà des préoccupations éthiques quant à la privation d'accès à un dispositif à un public défavorisé, certains acteurs de la mise en œuvre l'ont vécu comme une remise en cause de leurs compétences et de leurs pratiques professionnelles d'analyse des besoins et d'orientation des bénéficiaires. Ce pouvoir d'interprétation de la norme et de décision d'accès à la prestation des « *street level-bureaucrats* » - des acteurs de la mise en œuvre au contact des bénéficiaires -- est souligné depuis les années quatre-vingt par les travaux français et anglo-saxons de science politique⁸⁸³. Il s'avère particulièrement structurant pour les opérateurs de l'action sociale ou éducative. Le porteur du projet « APIPAQ » développe : *« Il y a eu une levée de boucliers des conseillers. (...) En tant que prescripteurs ils ne pouvaient pas orienter et faire jouer la chance ou le hasard, c'était incompatible, dans la prise de conscience d'un professionnel, de considérer que « si j'oriente, je ne suis pas sûr que le jeune en bénéficiera. Ça, ce n'est pas dans ma pratique. Si j'oriente, c'est que j'ai des compétences pour apprécier, évaluer l'attente du public et si j'oriente vers tel dispositif, je joue mon rôle de professionnel et j'attends que l'offre de service se réalise ». Il y avait un décalage avec ce que disaient les chercheurs sur l'intérêt de la randomisation et le professionnel au quotidien avec le public (...). Il y avait quelque part un sentiment d'injustice, de grand n'importe quoi, on était opposé entre l'approche scientifique et l'approche du terrain des professionnels »*. Cette réaction est d'autant plus forte si le dispositif propose un « traitement » important, par exemple les résistances ont été fortes pour le dispositif « APIPAQ » parce que *« les acteurs misaient beaucoup sur ce projet »*, alors qu'elles ont été moins présentes pour la « Mallette » parce que, pour reprendre les termes d'un évaluateur, *« on n'offrait pas aux classes bénéficiaires quelque chose de fou »*. Cette résistance fut aussi observée chez les agents de l'Anpe, dans le cadre de l'évaluation du dispositif « OPP/CVE ». Cette expérimentation organisait la diffusion d'un nouvel outil systématique de calcul

⁸⁸³ Michael LIPSKY, *Street-level bureaucracy: dilemmas of the individual in public services*, New York, Russell Sage Foundation, 2010; Vincent DUBOIS, *La vie au guichet*, Paris, Économica, 1999; Vincent DUBOIS, « Introduction, The sociology of administrative work, a study of "street-level bureaucracy" à la française », dans Jean-Michel EYMERI-DOUZAN et Geert BOUCKAERT, *La France et ses administrations: un état des savoirs*, Bruxelles, Bruylant, 2013, p. 169-177.

prévisionnel de la probabilité de sortie du chômage des demandeurs d'emploi, les établissant comme chômeurs potentiellement de longue durée et leur ouvrant accès à l'expérimentation. Cette informatisation permettant de maîtriser la subjectivité de l'orientation vers le dispositif a été très mal vécue par les conseillers puisque cette tâche était avant sujette à leur expertise professionnelle. Cette atteinte à l'expertise de jugement des acteurs de mise en œuvre se retrouve aussi dans l'Education nationale, comme l'explique un des porteurs de projets de l'internat d'excellence de Sourdun : *« on voit que ça les heurtait frontalement, car il fallait tirer au sort, et c'était très rapide, en deux heures c'était fait, alors que sinon ça nous aurait pris deux jours, et je me souviens d'un inspecteur d'Académie en fin de carrière qui me disait « je ne comprends pas ce que vous faites, je tiens des comités de sélection, si on se donne autant de peine c'est pour être aussi fin possible dans le choix des bénéficiaires. Ce que vous faites est anti-éthique, anti-professionnel »*. La nécessité d'orienter de potentiels bénéficiaires vers le groupe témoin est d'autant plus problématique lorsque ces derniers sont des publics très défavorisés. Les expérimentations où la population mère doit être préalablement mobilisée par les conseillers pour accepter de participer à l'expérimentation les amènent à un travail contradictoire : d'une part mobiliser le jeune en lui expliquant les atouts de ce dispositif et ensuite, s'il est sélectionné pour le groupe témoin, lui en refuser l'accès. Cette pratique est résumée par un porteur de projet de l'académie de Créteil : *« c'était compliqué de dire « venez, il y a un bon dispositif ! » et « ah... pas de chance, vous êtes dans le groupe placebo »*. On l'a assumé, mais ce n'était pas évident. »

Ces atteintes aux pratiques professionnelles des acteurs de terrain les ont menés à s'engager dans plusieurs types de travail institutionnel de déstabilisation de l'institution scientifique du tirage aléatoire. Un premier travail de déstabilisation intervient lorsque la population mère est élaborée sur la base d'un travail d'identification et de mobilisation des publics potentiellement bénéficiaires par des intermédiaires de terrain. Plusieurs cas d'expérimentations ont vu leur protocole scientifique pâtir d'une diminution de leur population mère à la suite du travail institutionnel déstabilisateur de ces intermédiaires.

Par exemple, dans l'expérimentation du CV anonyme, les conseillers de Pôle Emploi proposaient ce dispositif à des entreprises qui avaient le choix de l'accepter ou non et qui, lorsqu'elles donnaient leur accord pour participer, étaient ensuite tirées au sort. Or les conseillers qui avaient élaboré des relations privilégiées avec les entreprises étaient réfractaires à endommager celles-ci en imposant une contrainte à ces partenaires par la participation à l'expérimentation⁸⁸⁴. Les évaluateurs ont donc simplifié les procédures et centralisé le tirage au sort au niveau de la direction générale de Pôle emploi, un pilotage rapproché des agents a été organisé et *« pour remédier au faible taux d'acceptation des entreprises, la communication en direction des entreprises a été aménagée. Désormais, la participation à l'expérimentation n'est plus présentée comme un choix. Le conseiller informe le recruteur qu'il fait partie de l'expérimentation selon les termes suivants : “Nous vous informons qu'actuellement sur notre territoire, nous expérimentons le CV anonyme. La cible concerne les entreprises ayant un effectif d'au moins 50 salariés. Vous faites partie de cette cible. À ce titre, et pour ce recrutement uniquement, vous êtes susceptible de recevoir des candidatures sous format anonyme(...)»*⁸⁸⁵.

⁸⁸⁴ Un interlocuteur de Pole Emploi explique : « Il fallait mobiliser des outils de cv à distance et la direction le voyait comme quelque chose de simple à mobiliser pour eux et bien pris en main, mais dans les agences locales ils ne l'utilisaient pas. L'acteur clé ici c'était le conseiller donc il faut le convaincre que pousser l'entreprise dans l'expérimentation ça n'allait pas abimer sa relation avec l'entreprise. Il faut que le protocole soit léger et que les outils soient simples. »

⁸⁸⁵ Luc BEHAGHEL, Bruno CREPON et Thomas LE BARBANCHON, *Evaluation de l'impact du CV anonyme, synthèse et rapport final*, PSE,CREST,J-Pal, 2011, p. 14.

Une autre modalité de ce travail de déstabilisation à ce niveau peut être liée à l'ancrage des pratiques professionnelles et aux difficultés des acteurs de terrain à se plier à la logique de l'expérimentation et à ne pas procéder à une pré-sélection de potentiels bénéficiaires pour qui l'intervention serait le plus profitable. Cette tendance a été observée par exemple dans le cadre de l'expérimentation « APIPAQ ». Les décrocheurs devaient être repérés par des observatoires des ruptures de l'Education nationale et l'évaluateur souhaitait que tous les cas lui soient transmis, mais ces observatoires ne leur ont transmis que les dossiers de décrocheurs pour lesquels ils pensaient que le traitement serait une bonne chose et ont effectué de fait une première sélection subjective en amont.

Un second type de travail de déstabilisation consiste en une contamination volontaire par les intermédiaires de terrain de la composition du groupe témoin et du groupe test. Cette stratégie a été en partie mobilisée par les acteurs de mise en œuvre pour contrecarrer un faible *take-up*, notamment dans le cadre d'une contractualisation avec un sous-traitant pour lequel chaque personne traitée représente un versement. L'évaluateur de l'expérimentation sur les RMistes explique : *« On était face à un problème de non-recours à l'aide sociale et pour réduire ce non-recours, ils [les opérateurs] détruisaient le protocole expérimental en piochant régulièrement dans le groupe témoin. De leur point de vue, il n'y avait pas d'autres solutions, sinon de toute façon les ateliers étaient vides et donc il n'y avait pas de traitement (...). À chaque remise dans l'emploi le Conseil général le payait. Son objectif à court terme était donc de remettre le plus de personnes dans l'emploi. Donc il a fallu vérifier qu'ils n'avaient pas facturé des gens dans le mauvais groupe. »*. Ce travail institutionnel de contamination du groupe témoin peut être aussi pratiqué par un travail supplémentaire sur le groupe témoin effectué par des acteurs trouvant injuste leur non-accession au traitement. On observe cette pratique dans le dispositif « tous en stage », le porteur du projet explique : *« à l'inverse de nombreuses expérimentations que nous avons observées indirectement, nous n'avons pas renoncé à l'évaluation randomisée, ce qui a entraîné une sensation d'injustice dans la quasi-totalité des collèges, qui a entraîné des complexités d'organisation, et qui a encouragé la majorité des établissements à donner moins d'attention aux élèves suivis par l'association pour concentrer leurs efforts sur les élèves du groupe témoin »*⁸⁸⁶. On retrouve cette même pratique pour le dispositif AgiLoJe⁸⁸⁷.

Ainsi, lors de la phase de mise en œuvre, l'application pratique de l'assignation aléatoire révèle plusieurs oppositions liées au processus d'apprentissage imposé par cette institution scientifique. Pour les chercheurs, le respect du protocole scientifique constitue le cheminement nécessaire de l'apprentissage scientifique expérimental qui de plus requiert de retirer toute intervention subjective pouvant altérer le jeu du hasard. Cet effacement des interventions subjectives va à l'encontre des pratiques des acteurs de terrain de l'action publique qui sont eux à l'inverse aux prises avec l'humain dans leur travail quotidien. Cette subjectivité est particulièrement menacée par la systématisation requise par les évaluateurs du diagnostic des publics et de leur orientation, touchant

⁸⁸⁶ REUSSIR MOI AUSSI, « Tous en Stage », Note de restitution finale du porteur de projet, Fonds d'expérimentation pour la jeunesse, 2011, p. 10.

⁸⁸⁷ On lit dans le rapport d'évaluation : *« Il est à noter que les jeunes témoins ont pour un certain nombre d'entre eux pu bénéficier d'actions réservées au départ aux seuls bénéficiaires. Les sites ont justifié ces pratiques par des considérations tenant à la faible mobilisation des bénéficiaires, qui ont souvent été peu nombreux à se présenter aux réunions d'information. Ces dernières ont donc été ouvertes aux non-bénéficiaires afin de rentabiliser ces actions Enfin, certains sites ont pu évoquer des raisons éthiques pour expliquer que les jeunes témoins soient accompagnés de la même manière que les bénéficiaires, comme l'ont expliqué certains des acteurs interrogés : « On ne peut pas laisser les jeunes de côté sous prétexte qu'ils sont témoins » « De quel droit, sous prétexte que la « bécane » [l'extranet qui permettait d'effectuer le tirage aléatoire] a dit qu'il était non bénéficiaire, je le priverais de ce service alors que maintenant on a les moyens de lui répondre ? »*. CRÉDOC, AgiLoJe, Rapport d'évaluation, Fonds d'expérimentations pour la jeunesse, 2012, p. 16.

ainsi au cœur de métier des acteurs de la mise en œuvre. Ensuite, la répartition aléatoire prive le projet d'une partie de ses bénéficiaires potentiels, et empêche souvent la promotion du dispositif par les porteurs. Cette contrainte peut mettre les acteurs de la mise en œuvre en difficulté pour conduire leur action, faute de participants. Enfin, l'expérimentation aléatoire présuppose un terrain vierge de pratiques déjà existantes liées au projet expérimental. La préexistence de plusieurs éléments du projet avant l'expérimentation formelle a constitué un obstacle à l'observation d'un vrai groupe témoin et a aussi pu être source d'un travail institutionnel important des chercheurs pour lutter contre ces pratiques déjà installées qui endommageaient le protocole expérimental.

3.2. Confrontation des types de travail institutionnel autour du « traitement »

Les exigences du tirage aléatoire sont souvent identifiées comme les principales contraintes imposées par l'évaluation par assignation aléatoire. Néanmoins, dans l'utilisation de cette méthode apparaît un deuxième type de contrôle, portant sur l'intervention en tant que telle, qui nous semble témoigner davantage de l'ingérence des chercheurs dans la conduite de l'action publique, ingérence caractéristique de cette méthode d'évaluation. Ce nouvel intérêt des économistes appliqués pour la maîtrise de l'intervention, dans le but de contrôler précisément ce qui est évalué, les amènent à se rapprocher des acteurs de la mise en œuvre, ce qui constitue une pratique nouvelle pour les chercheurs. Nous étudions dans un premier temps les dispositifs de contrôle déployés sur la nature du dispositif et nous revenons ensuite sur l'apprentissage par les évaluateurs du processus de mise en œuvre de l'action publique.

3.2.1. La recherche de la connaissance de la « boîte noire »

La particularité de la méthode expérimentale est de requérir un contrôle complet sur l'intervention qu'elle évalue afin de s'assurer que l'administration du « traitement » est bien la même, indépendamment du partenariat d'acteurs, du conseiller et du bénéficiaire, et de pouvoir prétendre à une généralisation des résultats par la suite. En effet, cette méthode estime que le traitement envisagé lors de la conception du projet doit rester stable pendant toute la mise en œuvre de ce dernier. Le protocole de collecte de données ne prévoyant pas une vérification des étapes de la mise en œuvre et des détails de cette dernière, elle doit s'assurer que cette « boîte noire » reste stable et composée du même contenu d'intervention. Ce contrôle du traitement apparaît à deux niveaux, soit dès la conception du projet lorsque les chercheurs ont participé activement à cette phase, soit lorsque le projet est déjà existant lors de la formalisation de ce dernier, dans le cadre de l'expérimentation. Néanmoins, dans le cas de pratiques déjà installées ce contrôle devient problématique, voire impossible.

3.2.1.1. Le contrôle scientifique du traitement en amont

Nous avons déjà établi précédemment, que dans plusieurs cas les chercheurs avaient fortement contribué à la formalisation des projets expérimentaux voire parfois en étaient les initiateurs. Au fur et à mesure des expérimentations, les chercheurs apprennent à cadrer davantage le traitement, d'une part pour s'assurer que ce dernier rencontre bien leur intérêt académique et, d'autre part, pour s'assurer des composantes du « traitement » dont ils estiment l'impact.

Les évaluateurs peuvent cadrer l'intervention en imposant une formalisation de celle-ci. Lorsque l'intervention était nouvelle, les chercheurs ont contribué à élaborer un

« traitement » standardisé permettant de limiter les variabilités possibles sur les différents sites de mise en œuvre. Par exemple, lors d'incitation par courrier à participer au traitement, ils ont contribué à formater ceux-ci afin qu'ils soient identiques et ils ont établi avec les porteurs de projets des kits guidant les modalités d'intervention pour les agents de la mise en œuvre. Dans le cas de l'expérimentation AQ3E conçue par les chercheurs, ces derniers ont élaboré un « véritable outil « clé en main » et directement transposable dans n'importe quel autre établissement d'enseignement supérieur »⁸⁸⁸. Dans le cas de la « Mallette des parents », le terme même de « mallette » souligne cette standardisation. Pour cette expérimentation ils ont mis en place un programme de réunions ainsi qu'un DVD commun pour tous les établissements. Lors de la deuxième année de traitement, c'est eux qui ont décidé de réorienter l'expérimentation sur les procédés de sollicitation (*outreach*) de la participation des parents en formalisant les SMS, les DVD et les lettres de sollicitations qui sont exposées dans le rapport final⁸⁸⁹. Le porteur de projet évoque ce travail de formalisation : « C'était intéressant, ça nous a obligés à préciser et affiner ce qu'on devait faire, il y avait une tendance à dire « on va faire ça » et puis...l'évaluateur était très exigeant sur ces sujets là. Il disait « il faut un dispositif qui nous permette de dire si ça a marché, qu'est-ce qui exactement a marché », donc il a été très associé à ce travail là, donc il était du côté de l'évaluation et de la conception. (...) C'est comme ça qu'on a conçu le kit « mallette des parents » et c'est un peu la première fois qu'on a un dispositif qui fait l'objet d'un kit, sûrement à cause de l'évaluation, avec une dimension communication, avec logo, avec DVD sur le dispositif en lui-même. On n'aurait pas été aussi fin sur la façon de le présenter et de le mettre en forme sans eux ». Les évaluateurs apprécient aussi ce pouvoir d'ingérence, comme en témoigne un évaluateur de l'École d'Économie de Paris : « Là on a l'occasion d'expérimenter, de ne pas être simplement évaluateur, de choisir des choses, de poser des variantes, d'interagir directement sur le terrain avec les gens qui mettent en œuvre. Voilà on est dans un rôle plus proactif, plus en amont. »

⁸⁸⁸ UNIVERSITE DU MAINE, *Note d'étape sur l'expérimentation AQ3E*, Mission d'animation du Fonds d'expérimentations pour la jeunesse, 31 décembre 2010.

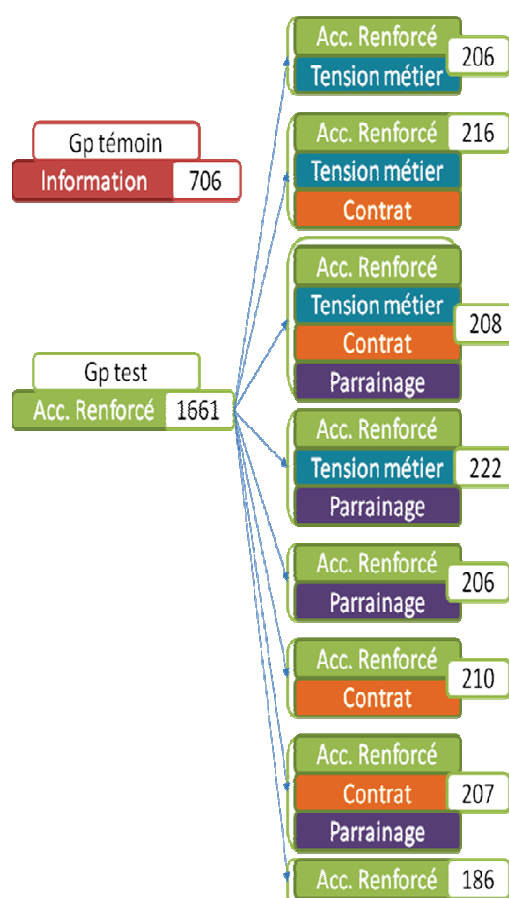
⁸⁸⁹ RECTORAT DE CRETEIL, *La Mallette des parents*, APDIESES-11, *Note de restitution finale remise par le Rectorat de Créteil, porteur du projet, au Fonds d'Expérimentations pour la Jeunesse dans le cadre de l'appel à projets lancé en 2008 par le Ministère en charge de la Jeunesse*, Rapport d'évaluation, 2011, p. 68.

Le projet Rebonds est emblématique de cette recherche de manipulation expérimentale et de contrôle du traitement⁸⁹⁰. La note d'étape sur l'évaluation rédigée sous forme d'hypothèses de recherche stipule en effet que cette expérimentation a pour but d'affiner les résultats des expérimentations précédentes et de comprendre la « boîte noire » de l'accompagnement renforcé⁸⁹¹. Un des évaluateurs explique : *« c'est nous qui avons initié ce truc-là. Car sur OPP/CVE on voit qu'il y a un effet de l'accompagnement renforcé, mais à côté il n'est pas non plus renversant. On s'est dit que ce serait bien que l'on puisse dire c'est quoi un accompagnement »*. Les chercheurs ont été ici à l'initiative de la démarche de contractualisation, ainsi que celle de l'orientation vers les métiers en tension⁸⁹².

Cette expérimentation est l'occasion pour les chercheurs d'initier une intervention différenciée extrêmement cadrée afin de maîtriser complètement la nature du traitement en composant huit groupes de traitement.

Le contrôle sur le traitement peut être soit facilité, soit compliqué par la sous-traitance de la mise en œuvre selon la précision de cette contractualisation. Dans le cas d'une contractualisation très précise, comme cela a été le cas pour l'expérimentation « Jeunes diplômés », les évaluateurs connaissent le contenu du traitement en détail et émettent l'hypothèse que ce dernier est suivi par les acteurs de la mise en œuvre. À l'inverse la délégation de service à un prestataire peut parfois contribuer à ajouter des intermédiaires à la mise en œuvre de l'expérimentation et à éloigner davantage l'action du porteur initial du projet et de l'évaluateur, notamment cas pour l'expérimentation

Figure 16: Les différents types d'interventions et nombre de bénéficiaires des neuf groupes de l'expérimentation Rebond



⁸⁹⁰ Cette démarche est restituée dans le rapport d'évaluation : « La discussion engagée a vite permis de comprendre qu'il ne serait pas très utile d'évaluer l'action de l'AFIJ en tant que telle prise comme une boîte noire, car la validité externe aurait été très faible et surtout difficile à apprécier. Les évaluateurs ont ainsi proposé d'isoler des composantes importantes de l'accompagnement et mises en avant par la littérature économique ». CREST, *Rebond: programme d'accompagnement des étudiants décrocheurs de l'enseignement supérieur*, Rapport d'évaluation, Fonds d'expérimentation pour la jeunesse, 2012, p. 29.

⁸⁹¹ En témoigne ce paragraphe : « Il existe depuis quelques années un courant assez dynamique de recherche sur l'efficacité de l'accompagnement renforcé à la recherche d'emploi. La France n'est d'ailleurs pas en reste sur ce domaine où plusieurs expérimentations contrôlées ont été menées ou sont en cours sur cette question (voir Behaghel et alii). Les principales conclusions empiriques convergent certes sur le fait que l'accompagnement renforcé accélère généralement le retour à l'emploi (voir Cockx, Dejemeppe ou Crépon, Dejemeppe, Gurgand), mais de nombreuses questions restent encore ouvertes. Parmi celles-ci, on peut en mentionner une qui nous paraît importante: Qu'est-ce qui fonctionne le mieux dans l'accompagnement? » CREST, *Note d'étape sur l'évaluation, REBOND, pour les étudiants décrocheurs de l'enseignement supérieur*, Mission d'animation du Fonds d'expérimentations pour la jeunesse, 31 décembre 2010.

⁸⁹² « L'introduction du contrat est une idée de l'équipe d'évaluation. Le contenu du contrat a été élaboré conjointement par le porteur de projet et l'équipe d'évaluation. Enfin, l'orientation vers les métiers en tension est une suggestion de l'équipe d'évaluation. L'AFIJ a adhéré immédiatement à cette idée et en a conçu elle-même le contenu. » CREST, *Rebond: programme d'accompagnement des étudiants décrocheurs de l'enseignement supérieur*, Rapport d'évaluation, Fonds d'expérimentation pour la jeunesse, 2012, p. 30.

« APIPAQ »⁸⁹³.

Ce contrôle sur le traitement peut aussi être effectué à un plus faible degré, même lorsque le projet préexistait à son expérimentation formelle. Le contrôle s'opère alors par une formalisation des procédures du projet précédent, comme par exemple, pour le projet porté par la Mission Locale de Beaune. On lit notamment dans la note d'étape rédigée par le porteur du projet que l'expérimentation « *normalise le travail des tuteurs dans les trois missions locales à Beaune* » par une mise à jour des outils d'intervention, des plaquettes et la définition commune de motifs de ruptures, ce qui permet « *d'avancer au même rythme avec une démarche partagée, selon la même méthode auprès des jeunes et des entreprises* »⁸⁹⁴. Cette formalisation du traitement a pour vocation de cadrer les pratiques des acteurs de terrain afin qu'elles soient similaires de site en site. C'est aussi un travail préparatoire à une possible généralisation du projet à d'autres contextes. À tout le moins, dans ces cas-là, les chercheurs ont contribué à formaliser les objectifs du projet, et à définir une population cible⁸⁹⁵.

Ainsi, le protocole d'expérimentation aléatoire, puisqu'il collecte de l'information uniquement sur la situation des bénéficiaires à l'entrée et à la sortie du dispositif, doit nécessairement contrôler le traitement et s'assurer que ce dernier reste stable afin d'en estimer l'effet. La recherche de connaissance des composantes de l'intervention proposée par le porteur du projet fait souvent place à l'intervention du chercheur pour formater l'action du porteur, la standardiser, afin qu'elle soit plus homogène et donc évaluable. Cette recherche du contrôle du traitement expérimental s'ajoute donc au premier contrôle effectué par l'assignation aléatoire des bénéficiaires. À l'instar de cette dernière, ce travail institutionnel de maintien de l'institution expérimentale a pu être source de travail institutionnel de déstabilisation chez les intermédiaires de terrains qui y voient une menace à leurs pratiques.

3.2.1.2. La quête de la connaissance de l'incarnation effective du projet

Lorsque les chercheurs n'ont pas réussi à contrôler le traitement de l'intervention en amont, s'ajoute alors un travail supplémentaire de recherche de connaissance du traitement afin de comprendre la nature de l'intervention. Lorsqu'au protocole expérimental par assignation aléatoire est ajoutée une méthode qualitative d'étude du jeu d'acteurs et des processus de mise en œuvre, cette dernière prend en charge ce travail. Cela a été le cas notamment pour l'expérimentation « OPP/CVE » dont la connaissance du traitement et de ses effets a été approfondie par des études de cas, une enquête auprès des bénéficiaires des projets et dont les équipes pluridisciplinaires ont permis une approche méthodologique qualitative et quantitative⁸⁹⁶. En l'absence de cette complémentarité méthodologique, les chercheurs tentent de mieux connaître l'intervention par leur interaction avec le porteur. La majorité des évaluateurs rencontrés

⁸⁹³ Le porteur du projet explique lors d'un entretien : « (l'organisme de formation) a développé tout ce qu'elle voulait : visite de cinéma d'art et d'essai, micro trottoir, atelier mosaïque, et des parents ont aussi participé. Il pouvait librement déployer toute forme d'action (...) C'était du cousu main ils se sont fait plaisir »

⁸⁹⁴ MISSION LOCALE DE BEAUNE, *Note d'étape sur l'expérimentation, Inscrire les contrats en alternance dans une logique de parcours sécurisé*, Mission d'animation du Fonds d'expérimentations pour la jeunesse, 31 décembre 2010.

⁸⁹⁵ L'évaluateur du projet Apipaq explique : « on a repris leur package et l'organisme de formation avait le même fonctionnement et proposait les mêmes solutions qu'avant à mettre en place, donc on a rien modifié, là où il y a eu un gros travail c'est sur l'explicitation des objectifs, rendus possible par cette évaluation. Les choses n'étaient pas formalisées de manières précises, (...) là on définit la population cible et déjà c'était instructif, on s'est dit « c'est quoi un décrocheur » » et rapport d'évaluation « En amont du projet, l'évaluateur a eu à cœur de réfléchir, avec les expérimentateurs, à la formalisation des notions et des objectifs du projet. Il a notamment incité les expérimentateurs à s'interroger sur la définition du « décrocheur », mettant à jour la très forte marge d'interprétation existant autour d'un terme partagé ». Dominique MAILLARD et al., « Evaluation d'un dispositif d'accompagnement de jeunes en rupture scolaire », *Net.Doc* 2012.97, 31.

⁸⁹⁶ Ce fut aussi le cas par exemple pour l'évaluation de l'expérimentation du projet Apipaq et AgiLoJe

expliquent spontanément que, contrairement aux autres méthodes quantitatives, celle-ci les amène à être directement avec les acteurs de terrain et à mieux comprendre l'intervention publique. Un évaluateur du Crest explique : *« Ce qui est vraiment très positif dans cette méthode ce n'est pas intrinsèque à la procédure de randomisation, mais au fait qu'on s'intéresse ex ante au programme avant qu'il soit mis en place, il y a une bien meilleure compréhension de la nature du traitement , une compréhension intuitive du jeu d'acteur, des enjeux des partenaires du projet, aussi une meilleure compréhension de la question, on contextualise mieux le problème (...). Ça c'est positif par rapport à plein d'autres évaluations que je fais qui sont ex post et où ce travail-là c'est souvent de la reconstruction »*. En effet les porteurs du projet de Tulle se souviennent aussi de l'importance de cette sollicitation : *« ils étaient très en demande de savoir comment ça se passait sur le terrain ils nous expliquaient que eux ils avaient une mission macro et travaillaient sur le chiffre et ils avaient besoin de savoir comment ça se passait au quotidien, qu'est ce qu'on faisait avec le jeune, c'est quoi leurs problématiques. Ils étaient très en demande de vivre ce qu'on vivait, ils venaient une journée, ils assistaient à des entretiens. »*. Cette recherche de connaissance détaillée du dispositif est d'autant plus difficile si cette dernière propose une variété d'interventions et il devient alors impossible pour les évaluateurs, s'ils n'ont pas cadré celles-ci en amont, de distinguer le poids de chaque élément dans l'effet produit, comme ce fut le cas pour l'internat de Sourdu⁸⁹⁷.

Un chercheur de l'École d'Économie de Paris explique : *« il faut être sûr qu'on sait ce qu'on évalue. Donc là, il y a un vrai travail de suivi de ce qu'est le projet, qui change pas trop tout le temps. Assez souvent, pour des raisons statistiques, on est obligé d'évaluer un programme sur un temps relativement long, pour avoir suffisamment de personnes dans l'échantillon, donc le programme peut évoluer au cours du temps. Il est important de pouvoir comprendre finalement ce programme moyen qu'on évalue, ou finalement ces différentes variantes du programme »*.

L'accès au terrain est organisé dans le cadre de la méthode aléatoire par la conduite d'enquêtes auprès des bénéficiaires et non bénéficiaires afin d'établir leur situation de départ et celle après l'intervention. La méthode expérimentale nécessite la production de données primaires, au sens où elle ne s'appuie pas sur des données déjà existantes, mais produit des données propres à l'effet d'intérêt. Elle organise de fait, une proximité au terrain pour les économistes appliqués, démarche nouvelle pour ces derniers. Un évaluateur du Crest explique : *« les gens qui font des expérimentations aléatoires, ce sont les plus proches des autres sciences sociales parce qu'ils vont parler aux acteurs. On est les seuls à aller parler aux acteurs, et je vois que les gens qui sont dans les autres sciences sociales le voient comme une concurrence car, naturellement, on marche sur leur territoire en ayant pas du tout les compétences d'acteurs qualitatifs (...). Mais du point de vue des économistes, on est vu comme les mecs qui vont se trainer dans la boue, on nous dit « c'est pas ça qu'il faut faire, ce qu'il faut faire c'est réfléchir aux grandes questions, c'est se prendre la tête, c'est pas aller trainer avec les conseillers des missions locales »*. Un évaluateur du Céreq renchérit : *« avec la méthode aléatoire, on est dans le concret ! »*.

Elle nécessite de fait une large équipe d'enquêteurs pour organiser la passation des questionnaires ou le recueil de ces derniers. C'est cette ingénierie qui explique

⁸⁹⁷ On lit dans le rapport : que le dispositif expérimental : *« ne permet pas de distinguer l'effet propre de chacun des éléments qui caractérisent cette scolarité »* ECOLE D'ECONOMIE DE PARIS, *Les effets de l'internat d'excellence de Sourdu sur les élèves bénéficiaires : résultats d'une expérience contrôlée*, Rapport d'évaluation, Fonds d'expérimentation pour la jeunesse, 2013, p. 13.

principalement le fort coût de cette méthode, vis-à-vis des autres méthodes d'évaluation. Cette gestion du personnel est une prestation proposée par le J-PAL qui met à disposition de ses membres des assistants de recherche formés pour conduire ce type de tâches. Quand les équipes d'évaluation ne bénéficient pas de ces ressources en interne, ils sous-traitent ce travail à des instituts de sondage comme TNS-Sofres par exemple pour le cas du projet des permis de conduire.

La conduite de l'expérimentation demande ainsi un contrôle du traitement et une bonne connaissance du terrain afin de s'assurer de quelle intervention exactement est estimé l'impact, et si aucun biais ne doit être pris en compte dans cette estimation. Néanmoins, lorsque la mise en œuvre ne respecte pas le protocole établi au départ, celui-ci peut être difficilement ajusté sans mettre en péril la production des résultats. Cette proximité avec la mise en œuvre s'avère être une expérience source de nouveaux apprentissages pour les chercheurs.

3.2.2. Le processus d'apprentissage de la mise en œuvre

« Evidemment on avait essayé de partir du monde tel qu'on pensait qu'il existait, mais on s'est rendu compte que le monde réel bien souvent est différent (...). Le random est quasiment impossible à réaliser parfaitement parce qu'en effet la réalité n'est jamais ce qui devrait être dans la théorie. La difficulté, c'est qu'en pratique et sur le terrain, les choses sont loin de se passer tel que le dit la théorie, voire même ce que disent les textes institutionnels. ». Ces propos d'un évaluateur du Céreq témoignent de l'apprentissage des chercheurs par leur participation à ces expérimentations. Ce travail de contrôle sur le traitement et cette recherche de connaissance sur les pratiques, s'il est accepté par les porteurs de projets, est susceptible de rencontrer la résistance des acteurs de la mise en œuvre, particulièrement s'il vient interférer avec des pratiques déjà installées dans ce champ de l'action publique. Cette résistance n'est pas le fruit d'un travail de déstabilisation volontaire, mais plutôt les effets d'un modèle d'apprentissage des acteurs de la mise en œuvre, inscrit dans l'ajustement de la mise en œuvre au cours de cette dernière. La littérature de l'analyse des politiques publiques souligne depuis les années soixante-dix que la mise en œuvre fait partie intégrante du processus de construction des politiques publiques et qu'elle est le fruit d'une succession de micro-décisions, de négociations, d'apprentissages qui en font un processus évolutif, chaotique et non linéaire dans la durée⁸⁹⁸. Peu à peu les évaluateurs de projets réalisent le décalage existant entre le projet planifié et son incarnation, ce qui fragilise leurs résultats.

3.2.2.1. Les pratiques de la mise en œuvre, un travail institutionnel de déstabilisation

Les évaluateurs ont rencontré plusieurs obstacles à une mise en œuvre du traitement correspondant à la planification originelle, menaçant ainsi la production de résultats. Ces obstacles ne correspondent pas cependant à des dysfonctionnements de la mise en œuvre, mais sont bien des éléments propres à la conduite de nouvelles interventions innovantes, tels qu'une lente montée en charge ou la résistance des acteurs au changement. Aussi, la méthode d'évaluation aléatoire est parfois déstabilisée par la mise en œuvre de projets déjà existants, pour lesquels des pratiques se sont déjà installées.

⁸⁹⁸ On peut d'ailleurs souligner que ces travaux ont émergé en réaction aux conduites des premières expérimentations évaluées par la méthode aléatoire aux Etats-Unis dans les années soixante-dix Jeffrey L PRESSMAN et Aaron B WILDAVSKY, *Implementation: op. cit.*.

D'abord, lorsque l'expérimentation organise une intervention nouvelle, un temps d'apprentissage est nécessaire aux acteurs de terrain afin de se l'approprier. Ce temps d'apprentissage est responsable d'une lente montée en charge du dispositif à ses débuts⁸⁹⁹. Dans les cas des expérimentations « AQ3E » et « Rebond », la mise en place effective du dispositif a pris respectivement un an pour la première, et cinq mois pour la seconde⁹⁰⁰. La montée en charge est d'autant plus problématique lorsque les agents de la mise en œuvre du projet sont de nouveaux entrants dans le *champ d'action stratégique* de l'action publique. Cette difficile prise en main de l'intervention s'est retrouvée notamment, dans les expérimentations de Pôle emploi organisant la nouvelle intervention des opérateurs privés ou dans l'expérimentation de l'association Actenses, dont l'intervention reposait sur le bénévolat de parrains qui ne se sont pas manifestés en nombre suffisant. On lit dans le rapport de cette expérimentation : *« En pratique, la difficulté à recruter des parrains a conduit à proposer des parrains à un sous-ensemble (aléatoire) d'élèves et à compléter l'intervention par des ateliers pour tous.(...) En deuxième année de programme, la pénurie de parrains s'est avérée encore plus importante »* 196 élèves de seconde sur 317 avaient un parrain à la fin de l'année de seconde pour la cohorte 2009 (soit 61%) et cette proportion est ensuite passée à 101 sur 237 (42%) pour la troisième cohorte. Aussi, dans la plupart des établissements, le programme effectivement mis en œuvre s'est écarté des principes : une proportion d'élèves a disposé du programme complet, tandis qu'une autre accédait seulement aux animations organisées par l'association »⁹⁰¹.

Cette difficulté peut aussi être due à un travail actif de résistance des acteurs de terrain au projet expérimental en particulier lorsque des acteurs intermédiaires, autres que l'institution porteuse du projet, interviennent dans la mise en œuvre. On retrouve cette pratique dans l'expérimentation « tous en stage » dans l'Essonne qui a introduit des acteurs du secteur privé dans les institutions scolaires, lesquels ont dans un premier temps pu rencontrer les résistances de l'équipe pédagogique. Ainsi, le porteur du projet connaît déjà cette difficulté lorsqu'il envisage les conditions de généralisation⁹⁰². L'évaluateur de ces projets se souvient : *« Les principaux freins sont venus de l'intérieur, souvent le directeur d'école était volontaire pour participer, mais l'équipe pédagogique pas forcément (...). Il y avait les dispositifs de coach du privé qui venait coacher les élèves dans la classe du prof, donc choc de culture complet, et pareil pour les conseillers d'orientation. Dès que l'expérimentation évaluée implique l'action d'un intermédiaire qui dispense d'une façon ou d'une autre du traitement, c'est lourd pour nous ! Par exemple, si le prof ne distribue pas le questionnaire, on a eu ça !, il a le droit. Des fois on en a appelé au directeur et c'est des questions très concrètes »*. Le manque d'enthousiasme des acteurs de terrain pour le dispositif a d'ailleurs amené à

⁸⁹⁹ Cette lente montée en charge n'est pas uniquement propre aux expérimentations par assignation aléatoire. Le Crédoc souligne cette tendance pour tous les projets expérimentaux qu'il a évalués Isa ALDEGHI, Pauline JAUNEAU et Lara MULLER, *Bilan méthodologique des évaluations conduites par le Crédoc dans le cadre du Fonds d'expérimentation pour la jeunesse*, 300, coll. Cahier de recherche, CRÉDOC, 2012, p. 27.

⁹⁰⁰ On lit dans la note d'étape de l'évaluation Rebond : *« (les) mois de décembre 2009 à mars 2010 s'apparentent davantage à une phase d'apprentissage de la part des chargés de mission, la montée en charge du dispositif ne débutant réellement qu'au mois d'avril 2010. »* CREST, *Note d'étape sur l'évaluation, REBOND, pour les étudiants décrocheurs de l'enseignement supérieur*, Mission d'animation du Fonds d'expérimentations pour la jeunesse, 31 décembre 2010; Loïc DU PARQUET et Pierre-Yves STEUNOU, *« Améliorer la Qualité des Emplois Exercés par les Etudiants » (AQ3E), note de restitution finale.*

⁹⁰¹ Luc BEHAGHEL, Vera CHIODI et Marc GURGAND, *Evaluation de l'impact du programme de parrainage d'aide à l'orientation de l'association Actenses, Rapport final, Rapport d'évaluation*, Institut des Politiques Publiques, J-PAL Europe, 2013, p. 11.

⁹⁰² *« Régulièrement, les établissements sont désignés par l'Académie et ne sont pas toujours entièrement volontaires pour participer. La vision des difficultés rencontrées par les élèves peut être radicalement différente dans l'établissement et au niveau académique. Il faudra alors anticiper les réticences ou scepticismes présents dans les établissements »* REUSSIR MOI AUSSI, *« Tous en Stage »*, Note de restitution finale du porteur de projet, Fonds d'expérimentation pour la jeunesse, 2011, p. 8.

l'abandon de l'expérimentation comme dans le cas de l'expérimentation de la Bourse aux projets de classe dite « cagnotte ».

Ce manque d'enthousiasme, aussi bien qu'un enthousiasme démesuré, menacent la scientificité de la production de résultats ainsi que leur validité externe⁹⁰³. Les évaluateurs doivent aussi parfois faire face aux aléas des contextes environnant l'expérimentation, comme notamment les grèves étudiantes pour les projets avec les Universités⁹⁰⁴.

Une autre difficulté apparaît lorsque le traitement ou des éléments du traitement préexistent à l'expérimentation, car les pratiques déjà développées peuvent interférer avec le cadrage souhaité par les évaluateurs. Les évaluateurs réalisent que les pratiques de terrain s'écartent parfois des textes de loi sur lesquels se sont appuyés les protocoles. Un évaluateur du Crest souligne ces difficultés qui ont mené à l'abandon d'une expérimentation : *« Le problème c'est que c'est un dispositif préexistant, c'est pas un dispositif expérimental (...), ça ouvre un tas de difficultés, c'est que les « process » sont déjà là. Comment modifier les « process » qui existent ? Et les « process » en plus ne sont pas géniaux... »*. Dans une autre expérimentation, celle portée par la Mission Locale de Beaune, les évaluateurs du projet se sont rendu compte de pratiques de détournement du cadre juridique, et ont noté qu'une partie des apprentis pouvait aller jusqu'à l'obtention de leur diplôme sans avoir de contrat d'apprentissage formellement enregistré⁹⁰⁵. En effet, les exigences de contrôle de la population de l'expérimentation requis par cette méthode amènent les évaluateurs à sonder la mise en œuvre de manière détaillée et à se rendre compte de tous ces ajustements de terrain. L'évaluateur de cette expérimentation reconnaît : *« Là, on a un hiatus, et il est certain que lorsque l'on voit tous les problèmes, on est vraiment dans de l'expérimentation qui n'est pas du tout contrôlée. Mais on peut se poser la question de : est-ce que franchement la réalité du monde de l'apprentissage dans sa complexité peut se plier à ce type de protocole d'évaluation ? C'est plutôt un niveau d'éclairage, mais si vraiment on éclaire d'en bas et dans le détail, on peut se dire « ouh là là tout ça c'est un peu du terrain mouvant » »*.

Enfin, une dernière difficulté réside dans la pratique des acteurs de terrain d'interpréter la norme et d'ajuster leur action au fur et à mesure de la mise en œuvre. Le traitement ne peut être figé pour toute la période de la mise en œuvre, ni homogène. Un porteur de l'expérimentation de Tulle explique : *« ils ont évalué deux groupes, mais toute chose n'était pas égale par ailleurs, chaque partenaire continuait à avancer, on a des conventions pluriannuelles d'objectifs qui nous obligent à avancer. Donc, comment continuer à faire une évaluation dans ces conditions-là ? »*. En outre, l'application du projet varie selon les sites au gré de l'interprétation des acteurs, notamment dans les dispositifs où le traitement repose sur l'intervention d'une personne. Dans

⁹⁰³ « Les questionnaires auprès des enseignants montrent une adhésion minoritaire au dispositif. Le résumé le plus simple est le suivant : moins d'un enseignant sur trois souhaiterait, si ce dispositif devait être expérimenté l'année prochaine, qu'une de ses classes y participe. 88% pensent que le dispositif ne s'attaque pas aux racines du problème, et 60% qu'il remet en cause les valeurs fondamentales de l'école. Par ailleurs, plus de deux enseignants sur trois pensent que le projet de classe lui-même est difficile à mettre en œuvre et constitue une charge importante pour les enseignants (...) Pour que l'évaluation soit valide scientifiquement, il faut que le fait d'évaluer ne fausse pas le fonctionnement du programme. Il faut pour cela que le dispositif soit suffisamment accepté pour que les acteurs ne modifient pas leur comportement en vue d'influencer les résultats de l'évaluation. » Luc BEHAGHEL et Marc GURGAND, Programme expérimental « Bourse aux projets de classe » : bilan de la phase pilote du point de vue de l'évaluateur, 2010.

⁹⁰⁴ L'évaluateur du projet conduit avec l'université de Marne la Vallée explique : « il y a eu des grèves étudiantes très importantes entre novembre et décembre de l'année où on mettait en œuvre le soutien et une partie du tutorat n'a pas été mis en œuvre parfois et donc le deuxième traitement a été imparfaitement implémenté. On n'est pas sûr de pouvoir aller au bout de l'évaluation de ce dispositif. Mais on pense qu'on aura de toute façon des choses à dire avec l'exploitation du premier volet sur l'orientation active »

⁹⁰⁵ Source : entretien avec évaluateur du projet : « On a vu des cas extrêmes, par exemple on a vu des élèves qui ont obtenu leur CAP, leur Bac Pro mais qui n'ont jamais eu de contrat enregistré »

l'expérimentation « Tous en stage », le porteur du projet souligne les différences de postures adoptées par les « coaches » intervenant en classe selon leur sollicitation par le professeur qui les orientait soit vers une intervention active avec proposition de stages, soit vers une position en retrait en guidant l'élève à prendre conscience de ses compétences⁹⁰⁶. Pour cette même expérimentation, le porteur souligne que la diversité des attentes envers et au sein d'une expérimentation peut mener à différentes interprétations des objectifs de celle-ci : *« L'expérimentation a été l'occasion de mieux comprendre comment gérer des attentes parfois radicalement différentes entre les objectifs de l'évaluation, les objectifs des co-financeurs, les objectifs des inspections académiques et la réalité des attentes des établissements, avec des positions parfois différentes entre le chef d'établissement, les enseignants et les élèves. »*⁹⁰⁷. Par ailleurs, plus le traitement nécessite une intervention dans la durée, plus il est sujet à devenir diffus, comme notamment dans le projet « Un pas dehors » évalué par le Crédoc, à cause notamment des variations et tâtonnements de la mise en œuvre concrète⁹⁰⁸.

3.2.2.2. L'apprentissage des pratiques de mise en œuvre par les évaluateurs

La conduite de la méthode d'expérimentation aléatoire, par le contrôle qu'elle requiert sur la mise en œuvre agit comme un révélateur de la plasticité de cette dernière. La proximité du terrain que demande cette méthode a été la source d'un apprentissage important des chercheurs la pratiquant, car aucune autre méthode d'évaluation n'avait autant rapproché les évaluateurs de la mise en œuvre. Beaucoup estiment, grâce à cette démarche, avoir amélioré la justesse de leur compréhension de l'intervention publique. Cependant, pour d'autres, ce rapprochement est aussi révélateur de l'impossibilité de saisir une intervention publique innovante dans son intégralité.

Cette interférence avec le terrain a été vécue comme une expérience riche en apprentissages pour les évaluateurs. Un évaluateur de l'École d'Économie de Paris explique : *« j'aime beaucoup le contact avec les gens sur le terrain qui concrètement mettent en œuvre les politiques , du coup je garde ma casquette quantitativiste et en même temps j'ai l'occasion même si je n'ai pas la prétention ou le temps de faire vraiment du travail qualitatif, mais j'ai un contact qu'on a pas autrement (...) à partir du moment où il y a un protocole, il doit être surveillé donc voilà. Du coup on a des espèces d'obligation de se rapprocher du terrain qu'on a pas forcément ailleurs »*. La méthode aléatoire nécessite aussi un recueil de données primaires qui est source de réflexion sur les biais liés à la conduite de questionnaires ou d'enquête⁹⁰⁹.

L'association avec les porteurs de projets ayant duré de deux à trois ans, les équipes de scientifiques qui entretenaient des relations expérimentales avec plusieurs porteurs de projets en même temps ont acquis rapidement des compétences d'interactions avec ces acteurs. Ils ont eu accès dans quelques cas aux enjeux financiers, politiques et sociaux environnant le dispositif et en ont apprécié la complexité. Après avoir conduit quelques

⁹⁰⁶ REUSSIR MOI AUSSI, « Tous en Stage », Note de restitution finale du porteur de projet, Fonds d'expérimentation pour la jeunesse, 2011, p. 9.

⁹⁰⁷ REUSSIR MOI AUSSI, « Tous en Stage », Note de restitution finale du porteur de projet, Fonds d'expérimentation pour la jeunesse, 2011, p. 9.

⁹⁰⁸ On lit dans le rapport d'évaluation : *« Les bénéficiaires n'ont pas toujours adhéré à l'accompagnement proposé et ont rarement pu être accompagnés de façon continue dans la durée. Dans les faits, l'accompagnement vers le logement a réellement pu être mis en place pour une minorité de jeunes, le suivi se limitant souvent à des appels téléphoniques répétés aux jeunes bénéficiaires : les référents interrogés ont souvent déclaré avoir rapidement « perdu le fil » avec des jeunes qui soit refusaient d'être accompagnés, soit avaient trouvé des solutions plus rapides ou plus efficaces par eux-mêmes ou par d'autres moyens »* CRÉDOC, AgiLoJe, Rapport d'évaluation, Fonds d'expérimentations pour la jeunesse, 2012, p. 6.

⁹⁰⁹ Un évaluateur du Céreq résume : *« Je crois que le plus grand intérêt de ces méthodes expérimentales est non pas le résultat c'est plus à nous faire réfléchir sur le protocole de recueil de données »*

expérimentations, ces compétences techniques apparaissent pour la plupart des évaluateurs comme la clé du succès d'une expérimentation aléatoire. Un évaluateur de l'École d'Économie de Paris revient sur cet apprentissage : *« C'est une espèce de compétence professionnelle qu'on a besoin d'acquérir, qui n'est pas classique dans la recherche. On apprend, assez vite à travailler avec les gens, comprendre leurs attentes, leurs difficultés et trouver des solutions pour rendre compatible un certain nombre d'objectifs qui sont ceux de l'évaluation d'un côté et ceux de l'opération de l'autre. Après chaque cas est particulier, on apprend rapidement que ce qui est important c'est ça. On apprend que les partenaires ont une logique différente qui a des aspects contradictoires, mais elle n'est pas moins légitime, elle est sur un autre plan. Après c'est des problèmes techniques où chacun fait des compromis, lâche un peu sur des points, c'est là l'essentiel de l'exercice, c'est le nœud qui fait que parfois on réussit ou pas. »*

Cette proximité avec les acteurs de terrain est aussi révélatrice pour certains des difficultés à comprendre le traitement, et amène certains scientifiques à remettre en question la pertinence des résultats avant même leur production pour cause de non-cadrage du traitement. Un évaluateur du Crest reconnaît : *« comment ça s'est réellement déroulé sur le terrain ? Je ne peux pas le dire. le protocole ça va il a été respecté, mais après comment ça se passait sur place je ne sais pas. (...) réellement ce qu'ils faisaient, je n'en sais rien. (...) le problème c'est qu'ils [les acteurs de la mise en œuvre] sont dans un monde ... moi j'ai besoin de savoir concrètement ce qu'on fait. Mais je ne l'ai jamais su »*. Ce problème est particulièrement présent lorsque le type d'intervention expérimenté est étalé dans le temps et basé sur une interaction humaine. Or ce type de traitement a été l'objet de la plupart des expérimentations de cette période qui portaient sur les accompagnements renforcés, les actions de tutorats, objets qui justement étaient difficilement préhensibles par d'autres méthodes d'évaluation. Le même évaluateur renchérit : *« un des problèmes gigantesques de l'expérimentation c'est que quand le dispositif n'est pas clair...quand vous donnez de l'argent à une entreprise pour recruter vous savez ce que vous faites, mais là vous ne savez pas ce que vous faites, vous payez quelqu'un qui va dire quelque chose (...) Le parrainage par exemple, on met des jeunes au contact de professionnels, on sait que ces jeunes vont les voir, une fois deux fois, qu'est ce qu'il s'est passé ? On ne sait pas, très difficile de savoir. Vous êtes dans l'incapacité totale de dire pourquoi il ne se passe rien. S'il se passe quelque chose, vous dites « ha oui on voit bien que le parrainage sert à ça »*

Ainsi, la nécessité de cadrage du « traitement » exigée par la méthode expérimentale par assignation aléatoire amène les économistes appliqués à se rapprocher du terrain. Paradoxalement, la pratique de cette méthode, quantitative par excellence, tend à faire émerger des préoccupations de nature qualitatives chez les évaluateurs. Ceux-ci prennent conscience de la plasticité de la mise en œuvre et de la distance entre le projet prévisionnel et son incarnation dans toute sa diversité.

Quels enseignements généraux pouvons-nous extraire de cette analyse de la phase de mise en œuvre de l'expérimentation aléatoire ? D'abord, nous avons identifié les limites de l'idéal de mise en œuvre sur laquelle se fonde cette méthodologie d'évaluation. Dans un protocole expérimental en science de la nature, le chercheur introduit lui-même le « traitement » source de l'effet qu'il souhaite observer sur son échantillon. Si un impact est observé, il sait donc exactement quelle en est la cause, notamment grâce au contrôle de toutes les autres variables pouvant constituer des sources alternatives d'effets, contrôle qui est garanti par la répartition aléatoire des groupes de comparaisons. L'idéal

scientifique du protocole expérimental demande donc un contrôle de la constitution des échantillons de comparaison (pour garantir l'interférence d'aucun autre facteur) ainsi qu'un contrôle du traitement (pour savoir quelle est la source de l'effet produit). L'application de cette méthode à l'action publique est fondée sur l'hypothèse que ces deux éléments peuvent être maîtrisés. Or, l'étude de ces incarnations empiriques révèle que cette institution nécessite un *travail institutionnel* de maintien conséquent des chercheurs pour que les expérimentations rencontrent ces pré-requis. En effet, la mise en œuvre des expérimentations dans le cadre de l'action publique nécessite l'intervention d'acteurs intermédiaires responsables de l'administration du traitement voire de la constitution des échantillons. Plus ces intermédiaires ont des marges de manœuvre, plus ils sont susceptibles de perturber l'intégrité du protocole. Ils peuvent aussi effectuer un *travail institutionnel* volontaire de déstabilisation lorsqu'ils jugent que ces exigences entrent en opposition avec leur éthique professionnelle ou leur savoir-faire professionnel dont notamment l'analyse de la situation des potentiels bénéficiaires et leur orientation. Ils peuvent aussi effectuer un *travail institutionnel* de déstabilisation uniquement par la conduite de leurs pratiques professionnelles habituelles de mise en œuvre qui par leur subjectivité et leur adaptabilité menacent l'homogénéité du traitement. Par ailleurs, la nature innovante des projets expérimentaux (dimension obligatoire, puisqu'une expérimentation doit déroger au droit commun pour refuser l'accès au traitement expérimental au groupe témoin) est un facteur de perturbation de l'application linéaire du protocole puisqu'elle implique des incertitudes sur l'adhésion des publics au dispositif (menaçant la taille des échantillons), et sur la mise en œuvre du nouveau traitement (le porteur de projet est-il en mesure de l'administrer correctement ? Toutes les difficultés potentielles de mises en œuvre ont-elles été anticipées ?).

Tous les cas d'expérimentations observés témoignent d'un décalage entre le protocole, les échantillons prévisionnels, ou le traitement envisagé par les chercheurs, et son incarnation sur le terrain. Les projets les plus stables sont ceux où les chercheurs sont fortement intervenus lors de la phase de conception pour cadrer l'intervention, la circonscrire et la standardiser, centraliser la procédure de tirage aléatoire et où les individus orientés vers le groupe test avaient quasi-systématiquement recours au dispositif proposé. En d'autres termes, des projets « objectivés », dénaturalisés laissant peu de place à la subjectivité des bénéficiaires et des opérateurs de terrain, et où ces derniers avaient peu de marge de manœuvre. A l'inverse, les projets les plus instables sont ceux conçus principalement par les porteurs de projets, faisant appel à un ensemble d'acteurs, s'inscrivant dans des pratiques professionnelles existantes, proposant une action modulable et étalée dans la durée et dont les opérateurs interviennent dans la constitution de la population mère (identification de potentiels bénéficiaires) et celle des groupes de comparaison. Ce difficile équilibre entre les intérêts de chacun questionne donc au final le maintien de la dimension clinique de cette relation, au sens où il semble que l'intérêt de l'une des parties prenantes n'est souvent atteint que par une diminution de l'intérêt de l'autre.

Dans les situations où les projets sont instables, le protocole expérimental est en difficulté pour appréhender par lui seul ces modifications. En effet, les hypothèses du contrôle expérimental induisent que ce protocole ne prévoit pas de dispositif de collecte de données pour comprendre quelle est la nature du traitement administré, puisque le caractère expérimental suggère justement la maîtrise de l'évènement perturbateur source du changement observé. Or les chercheurs/évaluateurs, par leur travail institutionnel de maintien du protocole, ont été amenés à se rapprocher du terrain, comme jamais auparavant, et ont pris conscience du décalage entre le projet prévisionnel et son incarnation. L'analyse des résultats, succédant à la phase de mise en œuvre devient alors

une tâche beaucoup moins linéaire et transparente que prévue, le processus d'apprentissage expérimental scientifique ayant été perturbé.

4. L'utilisation des résultats de l'expérimentation aléatoire par ses parties prenantes : la phase finale de la relation clinique

La dernière séquence de la relation clinique consiste en la production de résultats scientifiques et leur utilisation par les parties prenantes de l'expérimentation. La particularité d'une *relation clinique* réside dans la volonté des parties prenantes de l'expérimentation de satisfaire un besoin d'apprentissage par les résultats produits. Ce besoin est-il satisfait à la fin du processus expérimental ?

Nous nous intéressons, dans un premier temps, à la satisfaction des intérêts d'apprentissage des évaluateurs/chercheurs. La mise en œuvre des protocoles expérimentaux a compromis la production de résultats scientifiquement robustes et a complexifié l'analyse des données. Néanmoins, dans plusieurs cas, des résultats répondants aux attentes des chercheurs ont été produits grâce à cette méthode, ou encore à la mobilisation d'autres outils méthodologiques permettant de les renforcer. Nous avons cherché à comprendre quels résultats comportaient un intérêt pour les chercheurs et quelle utilisation académique ils en ont fait.

Dans un second temps, nous présentons l'utilisation des résultats par les acteurs de l'action publique. Celle-ci rencontre plusieurs obstacles. D'une part, leur disponibilité tardive compromet leur utilisation. D'autre part, l'objet de la connaissance produite par les chercheurs s'avère parfois peu propice à son utilisation autant *instrumentale* que *tactique*⁹¹⁰. À ce jour, seuls quelques cas témoignent d'une utilisation des résultats d'expérimentation aléatoire et nous cherchons à identifier les contextes favorables et défavorables à cette capitalisation pour l'action publique. Reste alors, la question ouverte de l'utilisation *conceptuelle*⁹¹¹ de ces résultats, c'est-à-dire son influence indirecte, cumulative, sur la conception des programmes publics par des acteurs extérieurs à cette relation clinique. La nature récente du phénomène étudié n'ouvre que sur des résultats provisoires à ce sujet.

4.1. L'utilisation scientifique des résultats de l'expérimentation aléatoire

Nous nous intéressons ici à l'atteinte des objectifs d'apprentissage des chercheurs s'étant engagés dans une *relation clinique* expérimentale. Cette dernière phase de l'expérimentation se découpe en deux temps. D'abord la phase de production des

⁹¹⁰ On peut identifier trois modèles d'utilisation des résultats d'évaluation : l'utilisation instrumentale ou rationnelle « *qui se manifeste par une influence directe d'un résultat de recherche sur une décision politique –le processus pouvant être tiré par la connaissance (une production de connaissance aboutit spontanément à une utilisation dans les politiques) ou par la résolution d'un problème (face à un problème, des décideurs commandent une recherche et en utilisent les résultats)* », Un modèle « conceptuel » ou « incrémental » où « *l'influence de la recherche se fait de manière indirecte et/ou cumulative, en apportant une connaissance qui modifie la façon dont les décideurs considèrent un problème et envisagent ses solutions* » et enfin « un modèle dit « *tactique* » ou « *symbolique* » (ou encore politique) dans lequel les décideurs utilisent, commandent ou sélectionnent des résultats de recherche pour défendre des politiques existantes ou supporter des décisions déjà prises et des changements souhaités » .Jocelyne DELARUE, Jean David NAUDET et Véronique SAUVAT, *Les évaluations sont-elles utiles? Revue de la littérature sur « connaissances et décisions »*, 03, coll. Série Notes méthodologiques Ex Post, Département de la Recherche, Division Evaluation et Capitalisation, Agence Française de Développement, 2009, p. 6.

⁹¹¹ Voir note précédente

résultats, durant laquelle le chercheur procède à l'analyse des données pour rédiger le rapport d'évaluation destiné au porteur de projet et au financeur. Ensuite la phase de capitalisation scientifique de ces résultats dans le monde de la recherche académique. Au regard de nos cas d'expérimentation, nous voyons que la phase de production des résultats ne correspond que rarement au modèle linéaire de comparaison de l'effet moyen sur le groupe témoin et le groupe test. Les complications de la mise en œuvre ainsi que les compromis avec le porteur de projets ont souvent perturbé l'analyse, laquelle appelle alors la consolidation par d'autres méthodes d'évaluation ou parfois, quand cette consolidation échoue, empêche la production de résultats rigoureux. Ces difficultés ont amené plusieurs chercheurs à modérer leur soutien à l'expérimentation aléatoire ou à en redéfinir le périmètre d'utilisation pertinente à la suite de leur expérience. Néanmoins, pour les cas ayant produit des résultats, la capitalisation scientifique est quasi systématique. Nous exposons alors ce que recherchent les chercheurs dans cet apprentissage final ainsi que le processus de transformation de ces résultats en matériel scientifique.

4.1.1. La production des résultats de l'expérimentation aléatoire

La méthode expérimentale par assignation aléatoire a été présentée par ses défenseurs comme une méthode d'évaluation dont l'analyse des résultats est transparente et linéaire et consiste simplement à la comparaison de l'effet moyen sur le groupe test et le groupe témoin. Cependant, au vu de nos différentes études de cas, il apparaît que la production des résultats a été beaucoup plus complexe, à cause notamment des aménagements et compromis effectués lors de la mise en œuvre du projet expérimental. Dans la quasi-totalité des cas, les évaluateurs ont rencontré des difficultés variées pour produire des estimations rigoureuses de l'impact de l'intervention étudiée. Une faible puissance statistique ainsi qu'une perméabilité des groupes de comparaison ont souvent engendré le recours à d'autres méthodes d'analyse statistique ou de collecte de données afin de renforcer la rigueur des résultats produits. Malgré ces efforts, beaucoup de rapports d'évaluation témoignent de cette difficulté à se prononcer sur la pertinence du dispositif au vu de l'impact d'intérêt.

4.1.1.1. Les difficultés rencontrées lors de la production des résultats

L'obstacle principal rencontré dans le processus de production de résultats est une faible puissance statistique des échantillons des groupes de comparaison constitués. Ce déficit numérique est dû, soit à un faible recours à la prestation, soit à l'attrition, en d'autres termes un manque de réponses aux enquêtes sur les individus constituant les groupes de comparaison. S'ajoute à cette difficulté la perméabilité des groupes de comparaison lorsque, comme nous l'avons vu précédemment, des individus du groupe témoin avaient accès au traitement ou des individus du groupe test n'y avaient au contraire pas accès. Ces deux difficultés engendrent la nécessité d'un effet visible et conséquent de l'intervention pour atteindre une significativité statistique, objectif difficilement atteignable pour des dispositifs de faible ampleur.

La réduction des échantillons par rapport à la planification originelle du projet est un problème rencontré par les évaluateurs dans la totalité des cas. Les difficultés rencontrées lors de la mise en œuvre ont toutes contribué à endommager la puissance statistique des échantillons. Le faible taux de recours au dispositif est la première source de diminution des échantillons prévisionnels. Celui-ci peut être conséquent dans des

expérimentations comme celle sur l'accompagnement renforcé des RMISTes, où seulement 17% des individus du groupe test ont recours au dispositif, réduisant un échantillon prévisionnel d'environ 13 000 individus à un échantillon effectif d'environ 2000⁹¹².

Cette baisse de puissance peut aussi être liée au faible taux de réponse des bénéficiaires aux enquêtes. Ce faible taux de réponse peut être la conséquence d'un travail institutionnel de déstabilisation des acteurs intermédiaires lorsque ces derniers étaient responsables de la collecte de questionnaires. Dans l'expérimentation « Tous en stage » par exemple, la non-distribution de questionnaires par les enseignants a contribué à restreindre l'analyse de données de 10 à 6 collèges⁹¹³. La diminution de la puissance statistique liée à l'attrition est d'autant plus importante lorsque l'expérimentation porte sur des publics fragiles, qui ont tendance à sortir des « radars » institutionnels, comme les « décrocheurs » scolaires ou les jeunes précaires. Cette difficulté est aussi rencontrée dans l'expérimentation « APIPAQ » où seulement 43% des décrocheurs ont répondu aux enquêtes finales⁹¹⁴. Or la réduction de la puissance statistique réduit la sensibilité de l'estimation de l'impact du dispositif. Dans une expérimentation aléatoire, moins les effectifs des échantillons sont importants, plus les effets de l'intervention doivent être élevés afin d'être considérés comme significatifs statistiquement (voir annexe 5). Par exemple pour le cas de l'internat d'excellence de Sourdun, à cause de faibles effectifs les effets de l'internat devaient atteindre *a minima* 30% d'écart type pour être considérés comme non issus de l'aléa statistique. Ceci signifie que *« si l'on fait par exemple l'hypothèse que 40% des élèves du groupe témoin obtiendront un baccalauréat général, il faudra que 55% des élèves du groupe test l'obtiennent pour qu'on puisse affirmer que cette différence est bien due au passage par l'internat d'excellence, et non pas à des aléas statistiques liés à l'analyse de populations relativement petites »*⁹¹⁵.

Le tableau ci-dessous propose de revenir sur la réduction des échantillons sur l'ensemble des cas étudiés, il présente la taille des échantillons effectivement assignés, le taux d'acceptation de cette allocation par les individus autant pour le groupe témoin que le groupe test, et la taille de l'échantillon effectivement suivi par le système d'enquête.

⁹¹² On lit sur la présentation de l'expérimentation sur le site du J-PAL : « Le taux de participation réel étant très faible, seulement 15%, la plupart des résultats sont imprécis » J-PAL, « Accompagnement renforcé des bénéficiaires du RMI, dans les Hauts de Seine, France », Poverty Action Lab.

⁹¹³ Le rapport d'évaluation explique : « (l'expérimentation a porté sur 10 collèges, 44 classes et 970 élèves. Néanmoins, les deux questionnaires n'ont été remplis de façon satisfaisante que dans 6 collèges. Dans certains établissements, les questionnaires n'ont pas été distribués dans les classes témoins. Dans d'autres, un questionnaire sur les deux était manquant ou les questionnaires n'ont pas été remontés à l'équipe d'évaluation. Le dépôt et la collecte des questionnaires s'effectuaient avec plusieurs mois d'intervalles et leur passation dans les classes nécessitait plusieurs intermédiaires. Cette complexité a été à l'origine de dysfonctionnements dans la chaîne de collecte qui s'est traduite par une diminution sensible du nombre d'observations » FEDERATION TRAVAIL EMPLOI ET POLITIQUES PUBLIQUES, *Tous en stage*, Rapport d'évaluation, Fonds d'expérimentation pour la jeunesse, 2012, p. 8.

⁹¹⁴ Source entretien avec l'évaluateur du projet

⁹¹⁵ ECOLE D'ECONOMIE DE PARIS, *Note d'étape sur l'évaluation des Internats d'Excellence*, Mission d'animation du Fonds d'expérimentations pour la jeunesse, 30 juin 2010.

Table 4: Présentation de la diminution des échantillons pour les 15 expérimentations étudiées

Expérimentation	Echantillon assigné	Echantillon acceptant l'assignation (Take up)	Echantillon suivi (attrition)
1. Accompagnement renforcé des bénéficiaires du RMI	13 236 test 4 430 témoin	17% du test (2 000 environ) 6% du témoin participe	Intégralité, pas d'enquête
2. jeunes diplômés	188 zones test (orientation à tx variable) total 31 854 individus 47 zones témoin	31% du groupe test (9890)	29 636 individus ont été tiré au sort pour être enquêté 21 431 ont répondu (78%). L'analyse s'est intéressée aux 11 806 sans emploi au moment de l'allocation.
3. OPP-CVE	45 000 affectés aux OPP 21% 72 000 au CVE 33% 101 000 parcours classique 46%	Environ 40% taux d'acceptation de l'orientation vers un accompagnement renforcé	Croisement données anpe et enquête au total 10 à 15% de l'échantillon n'a pu être suivi
4. Programme de prévention des ruptures dans l'apprentissage (Tulle)	Phase 1 : 768 test Phase 2 : 303 test 748 témoin 274 témoin	40% des jeunes du groupe témoin de la phase 1 ont eu un contact avec ces missions locales de même que 50% des jeunes du groupe témoin de la phase 2	Phase 1 : tx de réponse 84% Phase 2 : à 1 an 72% à 2 ans 56% Enquête tel à 2 ans : environ 72%
5. Mallette des parents 6 ^{ème}	Année 1 : 37 collèges 200 classes 100	1 réunion: 58% 2 réunions: 36% 3 réunions: 17% Formation 1: 11% Formation 2: 8% Formation 3: 3%	Suivi des fichiers élèves et conseil de classe
6. Actences	22 lycées 22 classes 630 élèves test 66 classes témoin	Parrainage : Au total 408 en classe de 2 ^{nde} , 180 en classe de 1 ^{ère} Ateliers : 94% année de 2 nd , 60% année de 1 ^{ère}	92% après la 2 ^{nde} 75% après la 1 ^{ère}
7. Internat de Sourdun	258 test 137 témoin	Test : 77% scolarisés après 1 an 63 scolarisés après 2 ans Témoin : 4% scolarisés après 1 an 9% scolarisés après 2 ans	Environ 90% d'enquêtés test et témoin
8. Tous en Stage	10 collèges au total 44 classes 970 élèves		6 collèges (28 classes) 550 élèves 271 test 279 témoin
9. AQ3E	234 test 234 témoin	Variabilité de l'exposition au traitement selon le nb d'offres d'emploi	Quasi totalité
10. Beaune	532 test 377 témoin 193 sélectionnés mais non entrés	55 individus du gp témoin ont eu accès au traitement	Impossible de calculer le taux d'attrition
11. CV anonyme	1 005 au départ	254 abandons Soit 385 témoin moyenne 57,2% 366 test	57,2% en moyenne
12. Rebond	706 Témoins 1661 test	94% accepte orientation vers l'acc. renforcé, 70% ont eu au moins un entretien, 91% orienté vers le contrat l'ont signé, 79% orienté vers le parrainage l'ont accepté et 45% ont rencontré un parrain	73% réponse enquête 1 63% réponse enquête 2
13. APIPAQ	513 test 511 témoin	149 soit 29% ont recours au traitement	43% test 42% témoin
14. 10 000 permis	5 370 test 1 795 témoin		Test : 1000 à 1 an, 1257 à 2 ans Objectifs 1 000 témoins, 828 à 1 an 629 à 2 ans
15. AgiLoJe	737 test 657 témoin	Environ 50% test n'a pas souhaité être aidé	69% à 1 an 58% à 2 ans

De fait, les chercheurs rencontrent des difficultés à établir strictement l'effet ou non du dispositif, notamment lorsqu'il s'agit d'interventions légères comme le Parrainage par exemple. Pour l'expérimentation « Actenses » on lit dans le rapport d'évaluation : « *les effets sont souvent à la limite de la significativité statistique (...), nous ne trouvons pas d'effet majeur et facilement interprétable sur l'orientation des élèves* »⁹¹⁶. La difficulté à établir l'effet du dispositif est aussi liée à un problème de temps de suivi des effets à

⁹¹⁶ Luc BEHAGHEL, Vera CHIOLDI et Marc GURGAND, *Evaluation de l'impact du programme de parrainage d'aide à l'orientation de l'association Actenses, Rapport final*, Rapport d'évaluation, Institut des Politiques Publiques, J-PAL Europe, 2013, p. 4.

moyens et longs termes du dispositif, notamment pour les dispositifs d'aide à l'orientation des jeunes, ce qui conduit les chercheurs à donner un statut partiel aux résultats, comme cela a été le cas pour cette même expérimentation⁹¹⁷.

Le nombre réduit d'individus dans chacun des groupes de comparaison accentue aussi les problèmes de comparabilité de ces derniers. En effet plus les groupes sont restreints, moins le tirage aléatoire permet de constituer des groupes identiques, et les risques d'échantillon aberrant augmentent. Cette différence entre les individus des groupes tests et témoins est vérifiée par un test de significativité à partir des données collectées sur les caractéristiques des individus composant la population des deux groupes. De larges échantillons, comme dans le cas de l'expérimentation des permis de conduire, permettent d'atteindre cette comparabilité des deux groupes, mais des effectifs réduits augmentent les risques de différences significatives entre les individus de chaque groupe⁹¹⁸. Cette difficulté a été rencontrée dans les expérimentations « Tous en stage », « Actenses » et l'internat de Sourdun⁹¹⁹. Les évaluateurs du dispositif de parrainage expliquent : *« De telles différences indiquent donc qu'en dépit du tirage au sort, les classes test présentent des caractéristiques spécifiques. Un tel cas de figure est possible : un tirage au sort malheureux sur un échantillon de petite taille (22 classes test seulement) peut conduire à un échantillon test particulier, surtout lorsque les classes sont hétérogènes (...). Au total, les différences entre le groupe témoin et le groupe test paraissent non négligeables. Elles sont suffisamment importantes pour suggérer qu'une simple comparaison de moyenne entre les deux groupes ne serait pas pleinement convaincante. Le risque est que les différences éventuelles qui seront relevées dans l'exploitation du deuxième questionnaire, après l'expérimentation, ne soient pas liées à l'effet causal de l'accompagnement, mais provoquées par des biais de composition »*.⁹²⁰

S'ajoutent à cette difficulté liée à la puissance statistique, les problèmes d'analyse de la significativité des résultats dans le cas d'une perméabilité des groupes test et témoin. Cette pratique que nous avons identifiée lors de la phase de mise en œuvre impacte fortement la pertinence des résultats. Cette difficulté est notamment présente dans les interventions proposant un dispositif composite proche du droit commun peu cadré par les évaluateurs en amont. Elle a été rencontrée par les évaluateurs du Crédoc dans l'expérimentation AgiLoJe où les « témoins » ont bénéficié du dispositif⁹²¹. Ainsi lorsque les évaluateurs de ce projet concluent à l'absence d'observation d'impact, ils précisent : *« Malgré l'existence d'un échantillon témoin randomisé, les effets spécifiques du dispositif n'ont pas pu être évalués dans le cadre de cette expérimentation du fait de la porosité du dispositif qui a semble-t-il été proposé aux bénéficiaires comme aux témoins. Cette absence d'impact de l'expérimentation peut s'expliquer principalement par le fait que les jeunes témoins ont souvent également bénéficié de l'accompagnement renforcé. (...) En conclusion, l'évaluateur ne peut se*

⁹¹⁷ Les évaluateurs écrivent dans le rapport : « on ne peut pas encore atteindre une précision statistique suffisante à des horizons plus longs. Ces résultats sont donc à prendre comme des résultats partiels » Luc BEHAGHEL, Vera CHIODI et Marc GURGAND, *Evaluation de l'impact du programme de parrainage d'aide à l'orientation de l'association Actenses, Rapport final*, Rapport d'évaluation, Institut des Politiques Publiques, J-PAL Europe, 2013, p. 18.

⁹¹⁸ TEPP-CNRS, *Faut-il subventionner le permis de conduire des jeunes en difficulté d'insertion? 10 000 permis pour réussir*, Rapport d'évaluation quantitative, Fonds d'expérimentation pour la jeunesse, 2012, p. 32.

⁹¹⁹ Ces différences sont décrites dans le rapport d'évaluation de « tous en stage » : « Les collégiens du groupe test déclarent moins fréquemment que leur père travaille et, lorsqu'il travaille, ils déclarent moins fréquemment sa profession. En outre, leur moyenne générale est un peu plus faible (de 0,65 point sur 20) et ils considèrent moins souvent que la meilleure filière pour réussir est la filière générale. Par ailleurs, ils portent un regard moins positif sur leur collège : moins d'élèves du groupe test considèrent que la discipline est respectée dans leur établissement et qu'il y a des bonnes relations entre élèves. » FEDERATION TRAVAIL EMPLOI ET POLITIQUES PUBLIQUES, *Tous en stage*, Rapport d'évaluation, Fonds d'expérimentation pour la jeunesse, 2012.

⁹²⁰ Luc BEHAGHEL, Vera CHIODI et Marc GURGAND, *Evaluation de l'impact du programme de parrainage d'aide à l'orientation de l'association Actenses, Rapport final*, Rapport d'évaluation, Institut des Politiques Publiques, J-PAL Europe, 2013, p. 20.

⁹²¹ CRÉDOC, *AgiLoJe*, Rapport d'évaluation, Fonds d'expérimentations pour la jeunesse, 2012, p. 4.

*prononcer sur l'effet propre du dispositif sur le parcours résidentiel et professionnel des jeunes. L'évaluateur ne peut pas se prononcer quant à la réplication du dispositif*⁹²². Cette difficulté d'analyse a été aussi constatée dans l'expérimentation portée par la Mission Locale de Tulle⁹²³.

Ainsi, beaucoup d'évaluations de ces expérimentations concluent à l'incapacité d'observer des effets significatifs de l'intervention. Il est alors difficile de distinguer si cette non-effectivité du dispositif tient uniquement au faible impact de ce dernier ou aux difficultés rencontrées par l'appareillage méthodologique pour l'apprécier, ou encore aux difficultés de mises en œuvre liées au protocole aléatoire. L'impact global du dispositif est estimé non significatif pour les expérimentations « jeunes diplômés », d'accompagnement renforcé des RMistes, « Rebond », « tous en stage » et, pour l'expérimentation « APIPAQ », l'évaluateur souligne même que l'effet nul est aussi lié au tirage aléatoire qui a handicapé le projet⁹²⁴. Beaucoup de résultats concluent à une effectivité mitigée du dispositif⁹²⁵. Comme l'expérimentation sur les permis de conduire qui démontre que l'expérimentation en facilite l'accès (objectif intermédiaire), mais n'augmente pas l'insertion des jeunes par ailleurs (objectif final)⁹²⁶. Dans cette même catégorie, on retrouve l'internat de Sourdun dont l'efficacité est significative uniquement sur les résultats en mathématique des élèves internes deux ans après leur entrée dans le dispositif, ou le CV anonyme qui s'avère corriger les biais d'homophilie (propension à favoriser à l'embauche les gens qui nous ressemblent), mais accentue les biais liés aux jeunes d'origine immigrée⁹²⁷.

⁹²² Ibid. p. 38.

⁹²³ On lit dans le rapport d'évaluation : « *Malgré une application rigoureuse du protocole de randomisation, 40 % des jeunes du groupe témoin de la phase 1 ont eu un contact avec ces missions locales de même que 50% des jeunes du groupe témoin de la phase 2. Ces contacts ont eu lieu à l'initiative des jeunes (ou incités par d'autres acteurs du monde de l'apprentissage). Les différences de traitement entre le groupe test et le groupe témoin sont assez faibles. De ce fait, le manque de puissance statistique rend difficile la quantification précise des effets de l'accompagnement* » CREST-DARES, *Sécurisation du parcours des jeunes s'engageant dans l'apprentissage : bilan d'une expérimentation aléatoire contrôlée conduite en Corrèze*, Rapport d'évaluation, Ministère des Sports, de la Jeunesse, de l'Éducation populaire et de la Vie associative, Fonds d'expérimentation pour la Jeunesse, 2012, p. 5.

⁹²⁴ On lit dans le rapport du projet Apipaq : « *Sans nier le fait que la méthode randomisée de l'évaluation quantitative a pu freiner l'adhésion de certains acteurs (membres des observatoires de rupture scolaire de l'Éducation nationale du conseiller de la mission locale), les résultats obtenus au regard des objectifs assignés induisent une préconisation de non-généralisation de l'expérimentation* » Dominique MAILLARD et al., « *Evaluation d'un dispositif d'accompagnement de jeunes en rupture scolaire* », Net.Doc 2012.97, 4.

⁹²⁵ On retrouve cette non-effectivité dans le rapport de l'expérimentation « tous en stage » : *Ce dispositif d'accompagnement n'a d'effet ni sur l'effectivité du stage, ni sur la qualité du stage, évalué au travers de la satisfaction du jeune, alors qu'il s'agit des objectifs visés par le dispositif. En revanche, il influence les choix d'orientation des élèves, ce qui correspond à l'objectif du stage lui-même. Les élèves accompagnés refusent moins fréquemment l'orientation vers des filières courtes et professionnalisées* FEDERATION TRAVAIL EMPLOI ET POLITIQUES PUBLIQUES, *Tous en stage*, Rapport d'évaluation, Fonds d'expérimentation pour la jeunesse, 2012.

⁹²⁶ Le rapport précise : « *Le taux d'obtention du permis de conduire est de 25,2% dans le groupe test à 12 mois et de 44,8% à 24 mois, contre 13,6% et 29,8% dans le groupe témoin. Il varie ainsi du simple au double entre les jeunes aidés et les autres au bout d'un an, avec une différence de 11 points. L'écart est encore plus élevé à 24 mois où il atteint 15 points. Cette différence est significative et peut donc être attribuée à l'impact causal de l'expérimentation qui produit effectivement un effet très net sur les chances d'obtenir le permis de conduire. Dans le cas de l'expérimentation, on ne constate d'effet notable de l'aide dont a bénéficié les jeunes, à court terme comme à long terme, ni sur les chances d'accéder à un emploi, ni sur la qualité des emplois occupés telle qu'elle peut être approchée au travers du montant du salaire, du type de contrat de travail, de la durée du travail, du statut d'emploi ou encore, du montant des revenus de remplacement* » TEPP-CNRS, *Faut-il subventionner le permis de conduire des jeunes en difficulté d'insertion ? 10 000 permis pour réussir*, Rapport d'évaluation quantitative, Fonds d'expérimentation pour la jeunesse, 2012, p. 8.

⁹²⁷ Luc BEHAGHEL, Clément DE CHAISEMARTIN, Axelle CHARPENTIER et Marc GURGAND, *Internats d'excellence: les enseignements de Sourdun*, op. cit.; Luc BEHAGHEL, Bruno CREPON et Thomas LE BARBANCHON, *Evaluation de l'impact du CV anonyme, synthèse et rapport final*, PSE,CREST,J-Pal, 2011, p. 2.

4.1.1.2. Une interprétation des résultats problématique appuyée par le recours à d'autres méthodes de collecte ou d'analyse

Ces résultats partiellement négatifs sont difficiles à interpréter pour les chercheurs qui n'ont souvent pas assez de clés avec leur protocole expérimental pour ouvrir précisément la « boîte noire » de l'intervention et comprendre les mécanismes à l'œuvre. Ils doivent alors *a posteriori* formuler des hypothèses interprétatives et prolonger la collecte de données par le recours à d'autres méthodes de collecte ou d'analyse statistique pour comprendre la cause de cet effet négatif. Par exemple le résultat surprenant de la non-effectivité du CV anonyme sur les candidats issus de l'immigration, résultat qui va à l'encontre des études de *testing* existantes à ce sujet, a amené les évaluateurs à prolonger l'analyse des données ainsi que l'interaction avec les opérateurs de Pôle emploi. Les évaluateurs prennent alors des précautions sur l'avancement de ces hypothèses : « *Reste l'hypothèse subtile, mais plausible d'une interaction entre signaux envoyés par le corps du CV et signaux envoyés par le bloc état civil : il se peut que l'anonymisation du CV, en ôtant de l'information sur les candidats, ait empêché les employeurs de réinterpréter à l'avantage des candidats potentiellement discriminés par les autres signaux du CV. Par exemple, les « trous » dans le CV pourraient être expliqués par un accès plus difficile à l'emploi lorsque le CV montre que le candidat réside en ZUS, mais pas lorsque cette information est masquée* »⁹²⁸. Le Parrainage aussi, dans le cadre du projet « Rebond », s'est avéré avoir un effet négatif sur le moral des jeunes. Bien que le protocole de cette expérimentation ait rigoureusement cadré l'intervention, il ne peut appréhender en détail la nature des interactions organisées par l'intervention entre le parrain et le bénéficiaire. Un chercheur du Crest explique : « *on a un peu décortiqué, on a ouvert un peu la boîte noire, mais on ne l'a pas complètement ouverte parce qu'on ne sait pas ce qu'il s'est passé avec les parrains. Là où il y a un effet négatif sur le moral, on ne sait pas trop. Donc à chaque fois il faut ouvrir des boîtes noires, et dès qu'on en ouvre une, il y a encore une nouvelle boîte noire qui arrive* ». Des interprétations sont alors possibles *a posteriori* par des échanges de nature qualitative avec les porteurs de projets. Ils émettent l'hypothèse que les entretiens avec le parrain ont eu pour effet de « *réduire la perception du champ des possibles et la confiance dans sa propre capacité à réussir son projet* », mais concluent qu'il est « *difficile de savoir ce qui s'est vraiment passé dans le cadre du parrainage* »⁹²⁹.

Ces difficultés d'interprétation de l'effet général de l'expérimentation poussent les chercheurs à étudier l'hétérogénéité des résultats et à tenter d'affiner ces derniers selon des catégories d'individus. Néanmoins, pour que cette pratique assure une rigueur des résultats, elle nécessite une puissance suffisante dans les sous catégories afin de les comparer entre elles. Un évaluateur de l'expérimentation « Jeunes Diplômés » revient sur cette démarche : « *c'est une époque où on avait un problème parce qu'on disait qu'il n'y avait pas d'effet externe (...), on se disait c'est bizarre, c'est vraiment bizarre parce que la théorie, quand même, elle est claire (...). C'est à partir d'un moment donné qu'on a commencé à identifier des effets externes, parce qu'on a commencé à restreindre les populations, on a regardé sur des populations très pertinentes* ». Cette démarche a été

⁹²⁸ Luc BEHAGHEL, Bruno CREPON et Thomas LE BARBANCHON, *Evaluation de l'impact du CV anonyme, synthèse et rapport final*, PSE,CREST,J-Pal, 2011.

⁹²⁹ CREST, *Rebond: programme d'accompagnement des étudiants décrocheurs de l'enseignement supérieur*, Rapport d'évaluation, Fonds d'expérimentation pour la jeunesse, 2012.

aussi choisie pour analyser les résultats non significatifs de l'expérimentation sur l'accompagnement des RMistes⁹³⁰.

Ces difficultés rencontrées lors de la production des résultats invitent à consolider les résultats produits par la méthode aléatoire à l'aide d'autres méthodes d'évaluation ou d'analyse de données. Ce recours à une méthode complémentaire est constatée dans l'intégralité des cas étudiés. Les évaluateurs peuvent compléter leur analyse soit par des méthodes qualitatives, soit par un traitement des données produites par d'autres méthodes statistiques permettant de corriger les hétérogénéités entre groupes ou de pallier à une faible puissance comme les régressions statistiques pour notamment vérifier le poids de certaines variables d'intérêt. Certains chercheurs avaient déjà prévu l'utilisation de méthodes mixtes, quantitatives et qualitatives, comme notamment les projets présentés par le Céreq et le Crédoc ou l'expérimentation « OPP/CVE ». Quand les résultats de la méthode aléatoire n'étaient pas assez robustes, ils ont alors renforcé la dimension qualitative de l'évaluation⁹³¹. On lit par exemple dans le rapport d'évaluation de l'expérimentation « Un pas dehors » : *« La montée en charge un peu plus faible que prévue du dispositif et l'attrition importante que l'on peut attendre parmi l'échantillon témoin montrant une faible adhésion au dispositif d'enquête, ont conduit à recentrer l'évaluation sur une démarche plus qualitative, en conservant pour les dernières vagues d'enquêtes, le seul échantillon bénéficiaire »*⁹³².

Une autre pratique courante mobilisée par les équipes d'évaluation ayant mis la méthode aléatoire au centre de leur démarche évaluative est de reconstituer deux groupes comparables par des méthodes d'appariement quasi expérimentales afin de résoudre les biais de sélection entre le groupe test et le groupe témoin. Cet outil de redressement a été notamment mobilisé pour les expérimentations « Tous en Stage », « APIPAQ », et de la Mission Locale de Beaune⁹³³, notamment lorsque les résultats étaient peu significatifs ou contradictoires avec d'autres études. Les chercheurs ont alors retravaillé les données avec des méthodes quantitatives non expérimentales. Comme dans le cas de l'expérimentation sur les RMistes ou le CV anonyme, un évaluateur du Crest explique : *« on essaie de raffiner la méthode pour voir si quand même on ne peut pas trouver un effet en utilisant d'autres variables explicatives qui n'étaient pas dans le modèle, pour diminuer l'imprécision des estimations (...), pour avoir quelque chose de plus précis et un effet qui ressort »*.

Lors d'un entretien, un évaluateur du Crédoc revient sur cette dénaturation du protocole d'expérimentation aléatoire, après avoir eu recours, pour exploiter les résultats, aux méthodes quasi expérimentales de matching de scores de propension et de différence des différences afin de réduire les biais de sélection des groupes. *« On peut dire « c'est très bien, vous n'êtes pas restés bloqués par rapport à un processus normé », et d'autres vont dire « attendez, il y a tellement d'adaptations, il y a tellement de rustines là-dedans »*

⁹³⁰ On lit dans un article de restitution: *“there is significant heterogeneity, and we can isolate a population of relatively recent welfare status, for which the program is more efficient, still not cost-effective from a strictly financial point of view, but for which the net budget cost is to the order of magnitude of the social value of employment.”* Bruno CRÉPON, Marc GURGAND, Thierry KAMIONKA et Laurent LEQUIEN, *Is Counseling Welfare Recipients Cost-Effective? Lessons from a Random Experiment*, p. 33.

⁹³¹ Isa ALDEGHI, Pauline JAUNEAU et Lara MULLER, *Bilan méthodologique des évaluations conduites par le Crédoc dans le cadre du Fonds d'expérimentation pour la jeunesse*, 300, coll. Cahier de recherche, CRÉDOC, 2012, p. 7.

⁹³² CRÉDOC, *Note d'étape sur l'évaluation « Un pas dehors »*, Mission d'animation du Fonds d'expérimentations pour la jeunesse, 31 décembre 2010.

⁹³³ On lit dans le rapport de l'expérimentation « tous en stage » : *« Pour remédier à ce risque de biais de sélection, nous avons reconstruit un groupe témoin avec la méthode du score canonique (ou « propensity score », en anglais) proposée par Rosenbaum et Rubin (1983, 1985) »* FEDERATION TRAVAIL EMPLOI ET POLITIQUES PUBLIQUES, *Tous en stage*, Rapport d'évaluation, Fonds d'expérimentation pour la jeunesse, 2012, p. 9.

que c'est aussi contestable, ou si ce n'est pas contestable, c'est difficilement probant ». Mais on va au fond des choses, ou alors, comme on fait classiquement dans l'analyse économique quand il y a de la poussière, on lève le tapis et on la met sous le tapis. »

Ainsi, les évaluateurs ont rencontré plusieurs types de difficultés qui ont contribué à complexifier la production de résultats par la méthode aléatoire uniquement, voire la compromettre. Néanmoins, ce processus constitua une source d'apprentissages académiques conséquents qui purent être valorisés dans les milieux scientifiques.

4.1.2. La capitalisation scientifique issue des expérimentations

La capitalisation scientifique issue des résultats est de plusieurs ordres et répond en partie aux objectifs d'apprentissages poursuivis par les chercheurs dans la pratique expérimentale. D'abord, ils ont acquis une expérience dans la conduite de cette méthode qui leur permet d'alimenter les réflexions académiques à ce sujet. Et ensuite, ils en extraient des données leur permettant de se positionner dans des débats internationaux en économie appliquée, ce qui transparaît dans les différentes publications académiques issues des expérimentations. Nous observerons, dans ce développement, les apprentissages différenciés des chercheurs selon les catégories d'appartenances identifiées précédemment (voir partie 2.1.1.1 de ce même chapitre).

4.1.2.1. L'apprentissage scientifique issu de la pratique d'expérimentation aléatoire

Les difficultés rencontrées lors de la production des résultats ont amené plusieurs évaluateurs/chercheurs à questionner la pertinence du statut d'étalon-or de la méthodologie d'évaluation aléatoire ou à produire des réflexions méthodologiques sur cette dernière.

Cette critique fut la plus forte chez les chercheurs appartenant au 3^{ème} groupe identifié précédemment, c'est-à-dire les *challengeurs*, dont le choix de la pratique de cette méthode émanait principalement d'une nécessité de se conformer à la hiérarchie méthodologie établie par le financeur. On retrouve donc cette remise en cause de l'expérimentation aléatoire comme « étalon-or » chez les chercheurs du Crédoc et du Céreq. Cette posture est résumée dans un rapport bilan du Crédoc sur sa participation aux expérimentations : *« L'expérience du CRÉDOC rejoint celle d'autres chercheurs, notamment le Céreq, qui a également évalué un nombre important d'expérimentations dans le cadre du FEJ. Les constats dressés remettent en question le dogme de la randomisation. (...) Difficile à mettre en œuvre en pratique, elle l'est tout particulièrement dans le cadre de projets de faible envergure qui concernent une population de taille réduite. En termes de méthodes, notre bilan nous incite à ne pas préconiser une utilisation exclusive des approches randomisées. L'approche qualitative a en effet permis d'apporter des informations précises et précieuses pour l'évaluation, notamment sur les questions de mise en œuvre des dispositifs et de fonctionnement des partenariats. (...) De façon plus générale, il importe de ne pas perdre de vue les autres méthodes d'évaluation et ne pas oublier que l'expérimentation reste une forme d'évaluation parmi d'autres. »*⁹³⁴ Certains de ces membres vont même jusqu'à questionner leur emploi dans le cadre des politiques sociales, un évaluateur du Crédoc explique lors d'un entretien : *« [cette méthode] permet de construire un échantillon*

⁹³⁴ Isa ALDEGHI, Pauline JAUNEAU et Lara MULLER, *Bilan méthodologique des évaluations conduites par le Crédoc dans le cadre du Fonds d'expérimentation pour la jeunesse*, 300, coll. Cahier de recherche, CRÉDOC, 2012, p. 64.

témoin incontestable, mais il reste plein de biais. Les politiques sociales c'est plus compliqué. Dans la passation des questionnaires et sur la mesure objective, on n'est pas sur le malade qui va mieux ou pas.(...) Il faut garder le principe de réalité, et ne pas tuer l'intervention sociale puisque l'évaluation est trop compliquée. Pour le RSA, on l'aurait tué dans l'œuf si on avait voulu faire du tirage au sort ».

Les chercheurs *titulaires* et *titulaires satellitaires* modèrent aussi leur propos en redéfinissant le périmètre d'utilisation pertinente de cette méthode. Au cours de nos entretiens et d'un retour sur leurs difficultés lors de la production des résultats, plusieurs promoteurs de cette méthode ont exprimé un apprentissage lié à la conduite de cette dernière qui les amène à questionner son recours systématique. Les évaluateurs du TEPP résument leur expérience en ces termes : « *On a expérimenté les expérimentations* » et « *L'évaluation aléatoire c'est très aléatoire* ». Un évaluateur du Crest développe ce propos : « *ceux qui sont convaincus disent que c'est super bien l'expérimentation, que l'on va faire que ça maintenant, alors qu'en fait ce n'est pas vrai. C'est super bien, certes, mais ça ne répond pas à toutes les questions. Ça a certaines limites, et si on ne présente que les bons côtés des choses, forcément, il y a des gens qui vont se braquer en disant « non ça ne répond pas à tout ». Il y a des problèmes qui existent quand même. Il faut un discours mesuré là dessus. Ce n'est pas forcément implémentable dans toutes les situations et ça ne peut pas répondre à toutes les questions non plus* ». Néanmoins, malgré ces difficultés, on retrouve dans les rapports d'évaluation des appels à la conduite de nouvelles expérimentations aléatoires sur les mêmes thèmes afin de consolider les résultats sur un type d'intervention. On retrouve cette requête dans plusieurs expérimentations comme par exemple celles portant sur l'accompagnement renforcé des RMistes⁹³⁵ dans le projet « Rebond »⁹³⁶, ou sur l'internat d'excellence de Sourdun⁹³⁷. Cette volonté s'est parfois incarnée directement par des appels à projets, comme ce fut le cas pour l'expérimentation d'accompagnement renforcé des jeunes en apprentissage de la Mission Locale de Tulle, laquelle fut prolongée à une échelle plus large par le Conseil National des Missions locales dans le cadre de l'appel à projet 1 du FEJ (AP1 112).

L'apport de résultats sur l'impact d'une intervention publique particulière n'est pas l'objectif d'apprentissage unique des chercheurs. Nous avons vu précédemment que ces derniers poursuivaient leurs propres intérêts de recherche dans la conduite de cette méthode d'évaluation, intérêts qui peuvent dépasser l'estimation de l'impact d'une intervention. Les évaluateurs rencontrés, notamment ceux du premier et deuxième groupe, sont en général plutôt satisfaits de cette expérience malgré la difficulté rencontrée à estimer l'impact. Nous proposons ici de revenir sur ces différents intérêts scientifiques satisfaits par l'expérimentation.

Afin de comprendre la démarche scientifique liée à l'expérimentation, il est avant tout indispensable de distinguer le travail de production du rapport d'évaluation, de celui de capitalisation scientifique. La capitalisation scientifique intervient à la suite de la

⁹³⁵ “Obviously, there is scope for more targeted intervention, and more experimentation would be needed to decide on which kind of population precisely” Bruno CRÉPON, Marc GURGAND, Thierry KAMIONKA et Laurent LEQUIEN, *Is Counseling Welfare Recipients Cost-Effective? Lessons from a Random Experiment*, p. 32.

⁹³⁶ CREST, *Rebond: programme d'accompagnement des étudiants décrocheurs de l'enseignement supérieur*, Rapport d'évaluation, Fonds d'expérimentation pour la jeunesse, 2012, p. 31.

⁹³⁷ « Si l'on souhaite porter un diagnostic plus large sur cette politique, il sera nécessaire de compléter notre travail par des études quantitatives portant sur d'autres internats d'excellence » Luc BEHAGHEL, Clément DE CHAISEMARTIN, Axelle CHARPENTIER et Marc GURGAND, *Internats d'excellence: les enseignements de Sourdun*, J-PAL Europe, Institut des Politiques Publiques, avril 2013.

production du rapport et consiste à se réapproprier les données collectées au regard des intérêts scientifiques poursuivis⁹³⁸.

Les évaluateurs du premier groupe expriment cette volonté de dépasser la seule estimation de l'impact d'une intervention dans un contexte donné. Un chercheur de l'École d'Économie de Paris explique qu'il s'intéresse particulièrement à des expérimentations permettant *« d'expérimenter des politiques très structurantes, des grands principes. Du point de vue de la connaissance scientifique, c'est ça qui nous intéresse. Ce n'est pas forcément que tel truc a marché comme ça, ici, maintenant. »* Un évaluateur du Crest renchérit : *« C'est typiquement une évaluation qui est pas vraiment faite pour évaluer si un programme marche ou pas, c'est une évaluation qui est faite pour dire finalement qu'est ce qu'il faudrait faire ? C'est quoi le problème ? C'est vraiment l'intérêt pour les mécanismes. »*. Ils s'inscrivent dans une utilisation conceptuelle des résultats, c'est-à-dire que *« l'influence de la recherche se fait de manière indirecte et/ou cumulative, en apportant une connaissance qui modifie la façon dont les décideurs considèrent un problème et envisagent ses solutions »*⁹³⁹. Ils s'intéressent donc à des mécanismes potentiellement structurants et communs à plusieurs interventions, comme notamment l'effet d'une contractualisation lors d'un parcours d'insertion. Ces intérêts sont guidés par des débats académiques existants, ou attirés par des mécanismes jusque-là inexplorés, comme par exemple l'activation du rôle des parents au collège dans l'expérimentation de la « Malette des parents »⁹⁴⁰. Cette recherche de connaissance ne correspond pas forcément aux besoins de connaissance des porteurs de projets, un évaluateur de l'École d'Économie de Paris revient sur cette même expérimentation : *« pour la mallette, sur le plan scientifique, [l'intérêt] n'est pas de dire : « les trois réunions telles qu'elles ont été menées, ça marche », ce qui est intéressant c'est de montrer que les parents sont un input [apport] dans la fonction de production scolaire et que l'école est capable d'agir sur les parents, et donc d'agir sur ce qui se passe en classe. On peut penser les parents comme un input scolaire et manipulable par l'institution scolaire. En sociologie de l'éducation, cette idée n'avait pas été démontrée. C'est ça qui m'intéresse dans une expérimentation comme ça. Après on peut parler de la validité externe, peut-être que dans certains cas ça ne marche pas, mais à la limite ça, ça ne m'intéresse pas beaucoup. Du point de vue de la connaissance, je veux mettre en évidence que ce mécanisme existe, donc qu'on peut y réfléchir et faire autre chose. (...). Moi je raisonne du point de vue de la recherche, c'est ça qui est intéressant. »*

En plus de mettre à jour de nouveaux mécanismes de l'action publique, les chercheurs souhaitent tester les limites de l'activation de ces derniers suivant une perspective cumulative de la connaissance, dépassant l'estimation de l'impact d'un dispositif en particulier. Un chercheur du Crest explique : *« Il y a un problème de communication [avec les porteurs de projets]. On n'a pas le même langage, ils n'attendent pas les mêmes choses. Chez nous les résultats c'est cumulatif, c'est-à-dire qu'il faut faire une synthèse (...), donc une étude elle-même ne clôt pas le truc et donc souvent, comme c'est étude par étude qu'on les présente, les gens se disent : « oui mais so what » [oui et alors ?] »*. Les évaluateurs sont donc dans une perspective d'utilité des résultats à long

⁹³⁸ Un évaluateur du Crest explique ce décalage en évoquant l'expérimentation portée par la Mission Locale de Tulle *« académiquement ce n'est pas fini d'un point de vu politique publique c'est relativement fini »*

⁹³⁹ Jocelyne DELARUE, Jean David NAUDET et Véronique SAUVAT, *Les évaluations sont-elles utiles? Revue de la littérature sur « connaissances et décisions »*, 03, coll. Série Notes méthodologiques Ex Post, Département de la Recherche, Division Evaluation et Capitalisation, Agence Française de Développement, 2009, p. 6.

⁹⁴⁰ Un évaluateur du projet explique : *« on y voyait un intérêt pour la littérature, un manque dans la littérature internationale qui pourrait aussi être comblé par une expérimentation de ce genre. »*

terme par les acteurs de l'action publique et une utilisation non restreinte aux seuls porteurs de projets.

Enfin, ils peuvent utiliser les produits dérivés des expérimentations grâce à l'ingénierie de sondage développée dans le cadre de ces dernières. Un des évaluateurs de l'expérimentation de Tulle explique : « *pour Tulle, la valorisation scientifique est plus indécise, on aura plus de mal à vendre l'article pour une revue, mais il y a un produit dérivé, c'est qu'on a eu un dispositif d'enquête très informatif sur la vie des apprentis en CFA et qu'on va pouvoir le mobiliser au-delà de l'expérimentation qui ne donne pas des résultats très forts* ». Ils peuvent aussi prolonger leur étude du dispositif par d'autres outils comme dans le cas de l'expérimentation sur les RMistes qui a été complétée par une analyse coût-efficacité notamment parce que ce type d'étude n'avait jamais fait l'objet de publications.

Ainsi, suite au rendu du rapport d'évaluation, l'utilisation de l'expérimentation aléatoire par les scientifiques se prolonge en dehors de la *relation clinique* expérimentale. Ces derniers transforment ces expérimentations en ressources propres au *champ d'action stratégique* de l'économie appliquée afin de conforter voire d'améliorer leur position dans ce dernier. Cette capitalisation, est particulièrement présente pour les chercheurs *titulaires* ou *titulaires satellitaires* et prend notamment la forme de publications scientifiques. Sa quasi absence chez les chercheurs *challengeurs* tient en partie à la différence de statut de ces derniers puisque les membres du Credoc et du Céreq (hors centres associés) sont moins soumis aux exigences du champ scientifique et des publications dans des revues académiques. Ils sont ingénieurs d'étude ou de recherche pour leurs institutions et ont seulement des exigences de publication dans ces revues institutionnelles et non pas dans les revues de renommée internationale.

4.1.2.2. La valorisation scientifique de l'expérimentation aléatoire

La valorisation scientifique prend beaucoup plus de temps que la production de résultats, par le travail supplémentaire qu'elle exige et par le délai de sélection par des revues à comité de lecture⁹⁴¹. Nous nous intéresserons principalement à ce type de publications qui constituent la principale ressource de positionnement dans le champ scientifique. Ce processus de capitalisation n'est pas finalisé pour tous les cas étudiés et notre recherche reste donc incomplète sur ce point. Cette valorisation scientifique était l'objectif premier des chercheurs *titulaires*, et *titulaires satellitaires* identifiés précédemment⁹⁴². L'atout de la méthode aléatoire est qu'elle permet d'inscrire ses praticiens dans les débats internationaux d'économie appliquée. Un chercheur du Crest explique lors d'un entretien : « *En termes de publication, c'est une opération qui est super rentable, on a des trucs qui sont vraiment bien, donc c'est ça que j'apprécie. C'est ça qui fait que je suis vraiment partant. On ne fait pas une évaluation pour faire une évaluation, ça n'a pas de sens.* » Ces publications académiques portent autant sur les réflexions méthodologiques issues des expérimentations que sur les nouvelles connaissances scientifiques produites. Beaucoup de ces papiers sont encore au stade de « document de travail » (*working papers*) et sont en cours de soumission à des revues. Ils sont pour la plupart écrits en anglais afin de favoriser leur diffusion internationale.

⁹⁴¹ Un évaluateur du TEPP et de l'Ecole d'Economie de Paris précise respectivement lors d'entretiens : « *en gros il faut deux ans pour avoir une bonne publication avec les révisions* » « *il reste encore un an et demi après pour écrire le papier* »

⁹⁴² Un évaluateur du Crest déclare : « *ce qu'ils [les évaluateurs d'expérimentation aléatoire] veulent faire c'est avoir une bonne publie dans une très bonne revue américaine* »

Plusieurs de ces papiers capitalisent sur les difficultés méthodologiques rencontrées lors de la conduite des expérimentations aléatoires en France. Ils développent notamment le recours à d'autres méthodes statistiques pour pallier aux biais d'attrition (déperdition du nombre d'enquêtés) comme l'indique la publication par l'équipe des évaluateurs du Crest dans les papiers de discussion de l'IZA (Institut pour l'étude du travail allemand), « Please Call Again: Correcting Non-Response Bias in Treatment Effect Models » (*Merci de rappeler : correction des biais liés à des non-réponses dans les modèles d'effet d'un traitement*)⁹⁴³.

Les valorisations issues de résultats d'expérimentations les plus prestigieuses dans des revues internationales d'économie sont à ce jour les résultats de l'expérimentation de la « mallette des parents » qui ont fait l'objet d'une publication dans une revue académique américaine, *The Review of Economics Studies*, en janvier 2014⁹⁴⁴; ainsi que les résultats de l'expérimentation « Jeunes diplômés » publiés dans la revue *The Quarterly Journal of Economics* en janvier 2013⁹⁴⁵. On note aussi à paraître la restitution de l'expérimentation de Pôle emploi sur les opérateurs privés de placement et le dispositif Cap vers l'entreprise dans *l'American Economic Journals: Applied Economics*⁹⁴⁶. Ces publications participent à une légitimation de cette démarche expérimentale *a posteriori* auprès du champ d'action stratégique de l'économie appliquée en France. Un chercheur ayant conduit des travaux expérimentaux explique : « Les autres membres du Crest ne sont pas finalement si convaincus que ça de l'expérimentation, ils le deviennent parce qu'ils voient qu'il y a beaucoup de papiers qui sont publiés. Pas par nous, parce que pour l'instant on en a pas beaucoup des papiers qui viennent de cette phase là, mais par les expérimentateurs américains. Ils se rendent compte que dans *QJE* (*Quarterly Journal of Economics*) ou *Econometrica* (*Journal of the Econometric Society*), il y en a pas mal des expérimentations. Alors ils se disent « ha oui, tiens, bon ok, force est de constater que les meilleures revues américaines [valorisent ces travaux] » ».

On relève aussi un ensemble de publications dans des revues académiques françaises. L'expérimentation « tous en stage » a fait l'objet de publications dans la *Revue Française d'Economie* en juillet 2013 ainsi que dans la revue *Education & Formation* en mars 2012⁹⁴⁷. Parfois la capitalisation s'effectue de manière croisée sur plusieurs expérimentations, ce qui est facilité par la conduite de ces dernières par les mêmes équipes d'évaluation ou les mêmes laboratoires. C'est le cas notamment des trois expérimentations sur le recours aux opérateurs privés de placement qui fait l'objet d'une publication dans la *Revue Française d'Economie*⁹⁴⁸. Les évaluateurs *challengeurs* valorisent plutôt leur expérience dans des publications en français dans leurs propres revues institutionnelles, n'ayant pas pour visée d'améliorer leur position, mais surtout

⁹⁴³ Luc BEHAGHEL, Bruno CREPON, Marc GURGAND et Thomas LE BARBANCHON, *Please Call Again: Correcting Non-Response Bias in Treatment Effect Models*, IZA 2012; Luc BEHAGHEL, Bruno CREPON, Marc GURGAND et Thomas LE BARBANCHON, *Sample attrition bias in randomized experiments*.

⁹⁴⁴ Francesco AVVISATI, Marc GURGAND, Nina GUYON et Eric MAURIN, « Getting Parents Involved: A Field Experiment in Deprived Schools », (2014) 81-1 *Review of Economic Studies* 57- 83, DOI : 10.1093/restud/rdt027.

⁹⁴⁵ Bruno CRÉPON, Esther DUFOLO, Marc GURGAND, Roland RATHELOT et Philippe ZAMORA, « Do Labor Market Policies have Displacement Effects? Evidence from a Clustered Randomized Experiment », (2013) 128-2 *The Quarterly Journal of Economics* 531-580.

⁹⁴⁶ Source: http://www.aeaweb.org/forthcoming/output/accepted_APP.php pour consulter le document de travail voir: Luc BEHAGHEL, Bruno CRÉPON et Marc GURGAND, *Private and Public Provision of Counseling to Job-Seekers: Evidence from a Large Controlled Experiment*, IZA.

⁹⁴⁷ Yannick L'HORTY, Emmanuel DUGUET et Pascale PETIT, « Une évaluation expérimentale d'un micro-programme social », (2012) Volume XXVII-1 *Revue française d'économie* 107-127; Yannick L'HORTY, Emmanuel DUGUET et Pascale PETIT, « L'accompagnement des collégiens dans leur recherche de stage: une évaluation expérimentale », *Education & formation* 2012.81.

⁹⁴⁸ Luc BEHAGHEL, Bruno CREPON, Marc GURGAND, Thierry KAMIONKA, Laurent LEQUIEN, Roland RATHELOT et Philippe ZAMORA, « L'accompagnement personnalisé des demandeurs d'emploi: les enseignements de trois expériences contrôlées en France », (2013) Volume XXVIII-1 *Revue française d'économie* 123- 158.

dans une ambition de réflexions sur leur pratique. Le Céreq et le Crédoc ont ainsi publié plusieurs retours d'expériences compilant leur participation pluriméthodologique à l'évaluation des expérimentations sociales évaluées par la méthode aléatoire ou non⁹⁴⁹.

Ainsi, la capacité des chercheurs à transformer leur participation aux expérimentations aléatoires en ressource mobilisable dans leur champ d'action stratégique s'avère variable selon leur catégorie d'appartenance. Les Chercheurs *titulaires* et *titulaires satellitaires*, qui publient ensemble puisqu'ils composent les mêmes équipes d'évaluation, et notamment ceux de l'Ecole d'Economie de Paris et du Crest affiliés au J-PAL, ont eu plus tendance à réussir dans cette tâche. Néanmoins, il est encore trop tôt pour faire l'inventaire de l'ensemble des valorisations scientifiques, car toutes ne sont pas encore publiées. A partir de l'analyse de nos entretiens, il semblerait que les intérêts d'apprentissage de ces deux groupes ont été atteints dans cette entreprise. La participation de ces chercheurs à une pluralité d'expérimentations a permis à chacun de conduire au moins une expérimentation ouvrant sur une valorisation scientifique. Mais qu'en est-il de l'utilisation des résultats par les porteurs de projets ? Ont-ils quant à eux réussi à capitaliser sur cette expérience ?

4.2. Une utilisation des résultats par les acteurs de l'action publique limitée.

La méthode expérimentale par assignation aléatoire a été promue comme une méthode qui, par la force de l'évidence de sa démonstration, favoriserait l'utilisation des résultats qu'elle produit par les acteurs de l'action publique⁹⁵⁰. Néanmoins, l'analyse des différents cas empiriques témoigne d'un ensemble de difficultés ayant contribué à une faible utilisation à ce jour par les porteurs de projets des résultats produits par cette méthode. Cette faible utilisation tient principalement à un décalage entre les impératifs scientifiques de la production de résultats par cette méthode et les intérêts d'apprentissage ou d'utilisation des acteurs de l'action publique. A l'issue de cette analyse, nous sommes en mesure d'identifier une utilisation des résultats des évaluations uniquement lorsque ces derniers vont dans le sens de l'intérêt des acteurs publics. Les résultats contradictoires sont eux laissés de côté voire délégitimés par les porteurs de projets ou les décideurs publics. Ces constats nous amènent à questionner la capacité de cette méthode d'évaluation à avoir été en France le véhicule d'une programmation politique ancrée sur des données probantes. Les promoteurs de cette méthode la présentent aussi comme un outil contribuant à nourrir, par une démarche cumulative, la décision publique sur le moyen et long terme. Faute de recul historique suffisant sur ce phénomène nous nous en tenons à des constats synchroniques sur ce phénomène.

4.2.1. Les obstacles à l'utilisation des résultats par les acteurs de l'action publique

Plusieurs éléments propres à l'expérimentation sociale voire à l'expérimentation évaluée par assignation aléatoire font obstacle à une utilisation des résultats par les porteurs de projets. On identifie d'abord des problèmes de temporalité se manifestant principalement par un long délai entre la fin de l'expérimentation et la production de résultats scientifiques. Un second élément obstruant l'utilisation des résultats est aussi

⁹⁴⁹ Isa ALDEGHI, Pauline JAUNEAU et Lara MULLER, *Bilan méthodologique des évaluations conduites par le Crédoc dans le cadre du Fonds d'expérimentation pour la jeunesse*, 300, coll. Cahier de recherche, CRÉDOC, 2012; Claudine ROMANI, *Le Céreq et l'évaluation des expérimentations sociales*, 84, coll. Net.Doc, Céreq, 2011.

⁹⁵⁰ Esther DUFLO, « Evaluer l'impact des programmes d'aide au développement: le rôle des évaluations par assignation aléatoire », *op. cit.*

un décalage qui se révèle lors de cette dernière phase de l'expérimentation entre l'intérêt d'apprentissage du porteur du projet et les résultats produits par l'évaluateur du projet. Ces éléments ont contribué dans plusieurs cas à une rupture de la *relation clinique* établie en amont de l'expérimentation.

4.2.1.1. Une temporalité de production de résultats scientifiques non adaptée aux besoins des porteurs de projets

Le premier obstacle à l'utilisation des résultats scientifiques pour éclairer l'action publique et informer sa nouvelle programmation est le décalage temporel entre le moment où le porteur de projets aurait besoin des résultats et le moment où l'évaluateur est en mesure de les produire. Cette période s'étend de 3 mois à 2 ans entre la fin de l'expérimentation et la production du rapport final. Cette période varie d'une part selon l'importance du traitement statistique à effectuer pour analyser des données difficiles à interpréter et d'autre part selon le temps nécessaire pour observer l'impact d'un dispositif après sa mise en œuvre. Ce long délai est source de rupture de la *relation clinique*. Néanmoins cette rupture au sein de la relation clinique n'est pas uniquement liée à une différente temporalité des acteurs mais aussi à un intérêt différent sur le contenu de l'évaluation nécessaire à un apprentissage. Ce décalage est donc aussi amplifié par les perceptions et les besoins d'information divergents des parties prenantes de l'expérimentation, correspondant à des pratiques différentes liées à leurs champs d'action stratégique respectifs.

Une des motivations premières du choix de cette méthode pour les porteurs de projets était de démontrer l'efficacité du dispositif afin de sécuriser leurs financements. Néanmoins, cette motivation n'a pas pu être satisfaite à cause de productions de résultats trop tardifs. Cette difficulté est liée à toutes les méthodes d'évaluation s'intéressant à l'impact d'une intervention, puisque cette opération nécessite de pouvoir observer l'évolution des bénéficiaires dans le temps. Elle est particulièrement problématique dans une démarche expérimentale puisque les porteurs de projets attendent la production de résultats pour légitimer leur intervention et pouvoir solliciter de nouveaux financeurs⁹⁵¹. Un évaluateur du Crédoc résume bien cette situation : *« le porteur réfléchissait à comment financer ça. Et ça a été problématique parce qu'ils avaient embauché, donc ils avaient besoin de trouver d'autres financeurs, et ils nous ont donc demandé des résultats pour aller démarcher des financeurs, mais nous on n'était pas en mesure de leur fournir des résultats à ce stade-là »*. Cette difficulté s'est ressentie dans toutes les expérimentations et particulièrement lorsque les porteurs appartenaient au groupe des *titulaires précaires* ou des *challengeurs* et donc avaient besoin de légitimer leur intervention pour renouveler leurs subventions. Un des porteurs de l'expérimentation de Tulle explique : *« il y eut un décalage fort entre la fin de notre action et l'évaluation. Nous, on a été pris par notre quotidien et on était déjà passé sur autre chose. Il y avait un décalage très fort entre l'évaluation et l'action. On a eu une réunion à la fin de l'action avec les financeurs locaux, car tous les acteurs voulaient continuer, les financeurs ont dit « nous on veut bien voir, mais on veut savoir les résultats ». Or on n'en avait pas, il y avait des résultats partiels, mais peu concluants »*.

⁹⁵¹ Cette dimension est évoquée dans un rapport du Crédoc « Concrètement, pour ces professionnels, fortement investis sur le terrain, la fin de la période d'expérimentation a signifié la fin des financements alloués aux projets, sans pour autant coïncider avec la date de diffusion des résultats » Isa ALDEGHI, Pauline JAUNEAU et Lara MULLER, *Bilan méthodologique des évaluations conduites par le Crédoc dans le cadre du Fonds d'expérimentation pour la jeunesse*, 300, coll. Cahier de recherche, CRÉDOC, 2012, p. 55.

Cette dépendance envers les financements du FEJ a mis plusieurs porteurs de projets en difficulté puisqu'ils n'avaient pas anticipé la fin des financements et imaginaient que si l'évaluation était positive le FEJ procéderait à une pérennisation du financement. Une interlocutrice du FEJ revient sur ce « malentendu » : *« je me suis retrouvée devant 700 personnes avec toutes les Missions Locales de France réunies en congrès en disant « ok et après on fait quoi ? c'est bien gentil, mais les financements vont s'arrêter. Qu'est-ce que vous comptez faire ? » Et je leur ai dit « je vous retourne la question : qu'est-ce que vous, vous comptez faire ? » »*. L'association « Réussir moi aussi » estime même que le financement du FEJ a eu l'effet pervers *« d'habituer »* leurs *« financeurs habituels à l'existence d'un cofinancier extérieur, qui nous ont demandé de présenter la même action après l'expérimentation, sans avoir le soutien financier du FEJ, ce qui était impossible pour notre budget. Les actions ont donc finalement pris fin sur l'ensemble des territoires concernés par l'expérimentation juste après, alors que l'action de l'association sur ces territoires avait commencé plusieurs années plus tôt, sans bien sûr que la qualité du dispositif soit remise en question. »*⁹⁵². Cette difficulté est principalement présente chez les associations, en témoigne la disparition de l'Afij, porteuse du projet « Rebond », qui cessa son activité en décembre 2013 faute de financements⁹⁵³.

Lorsque les porteurs de projets sont issus de la catégorie des *titulaires* du champ stratégique de l'action publique et que leur source de financement n'est pas menacée, ce sont les impératifs de la décision publique qui se retrouvent contrariés par ce décalage dans le temps. En effet, les expérimentations aléatoires ont aussi été conduites sur des interventions faisant polémique comme par exemple le recours aux opérateurs privés de placement dans l'accompagnement des demandeurs d'emploi, envers lesquelles les parties prenantes souhaitaient le plus rapidement disposer d'éléments de preuves pour appuyer leur argument. Un chercheur du Crest revient sur les attentes des décideurs : *« Ils auraient bien aimé avoir les résultats 15 jours après que ça ait commencé, même si l'objectif était de remettre les gens dans l'emploi. Il fallait expliquer que ça prend déjà six mois. J'exagère un peu, mais de toute façon c'est toujours trop long pour un homme politique. »*. Le temps du déploiement du protocole expérimental, de minimum deux ans, et ensuite le temps de la production de résultats, ont souvent été jugés trop longs par les porteurs de projets en attente de ceux-ci pour prendre une décision. En outre, la mise en agenda d'une nouvelle intervention publique s'inscrit souvent dans un contexte favorable, une fenêtre d'opportunité (*Policy window*) sur une courte période qui, lorsqu'elle est dépassée, compromet les possibilités de décision et d'adhésion au dispositif.

Malgré tout, le temps de l'expérimentation et de la production des résultats peut amener un changement d'interlocuteur de la *relation clinique*. Or, nous avons vu précédemment que ces *relations cliniques* et les compromis établis au sein de ces relations tenaient beaucoup aux parcours des individus les composant. Ce changement d'interlocuteur s'incarne souvent par un manque d'intérêt du nouvel interlocuteur pour les résultats de l'expérimentation qu'il n'a pas commandité. C'est ce qu'on observe pour le cas de la « Mallette des parents » – même si celle-ci a finalement bénéficié du soutien du premier

⁹⁵² REUSSIR MOI AUSSI, « *Tous en Stage* », Note de restitution finale du porteur de projet, Fonds d'expérimentation pour la jeunesse, 2011, p. 10.

⁹⁵³ <http://afij-actualites.org/2013/11/04/faute-de-soutien-de-letat-lafij-se-voit-oblige-de-mettre-un-terme-a-ses-activites-en-faveur-de-lemploi-des-jeunes/>

porteur du projet –, de l'accompagnement des RMistes et de l'expérimentation « Jeunes diplômés »⁹⁵⁴.

Il en résulte alors, une absence d'interaction finale entre évaluateur et porteur de projets autour des résultats. Un évaluateur explique pour l'expérimentation de l'accompagnement des RMistes : *« on ne savait pas à qui rendre le rapport à la fin de l'évaluation puisque les gens avaient changé de poste et on n'avait plus de contact »*. Ce manque d'intérêt est alors amplifié lorsque les résultats de l'évaluation du dispositif témoignent du peu d'effets de ce dernier. Un évaluateur du Crest revient sur le changement du cabinet commanditaire de l'expérimentation « Jeunes diplômés » : *« Alors là, ça a été décevant, car le commanditaire était le Cabinet, et le Cabinet avait changé, avec aucun souvenir de ça, ça c'est totalement décevant »*. Un interlocuteur de la Dares ayant servi d'intermédiaire sur cette expérimentation explique : *« ce programme a été lancé au départ au moment où le débat était assez sensible, ensuite c'est retombé un peu (...). Il y a eu des changements de gouvernement et après il n'y a pas eu un point de rendez-vous sur cette évaluation »*.

Ce décalage est accentué lorsque les évaluateurs rencontrent des difficultés lors de la phase de l'analyse des résultats et que le travail sur ces derniers s'avère plus fastidieux et long. En effet, les évaluateurs ont été réfractaires à émettre des recommandations pour l'action lorsqu'ils n'étaient pas satisfaits de la rigueur des résultats produits ou que ces derniers comportaient des ambiguïtés. Par exemple, dans le cas de l'expérimentation du CV anonyme, un des évaluateurs explique : *« si l'on repose la question de l'articulation avec la décision publique, [le problème] c'est que ce deuxième temps d'analyse où l'on tisse cette richesse supplémentaire des données, mais avec des hypothèses, prend du temps, et il vient après la publication des résultats publics. Donc, quand il y a incertitude sur l'interprétation, il y a urgence de ne pas se presser. Je crois qu'il aurait été utopique de se dire on va avoir six mois, un an supplémentaire pour faire des analyses supplémentaires, les résultats étaient attendus. Je pense qu'il y a une espèce de superposition du temps politique et scientifique, et c'est différent. »*

4.2.1.2. Les dérivations d'objet évaluatif d'intérêt

Le second obstacle à l'utilisation des résultats est issu du décalage qui survient au cours de la *relation clinique* entre l'intérêt de connaissance du chercheur et celui du porteur de projets. Cette distance entre l'objet de l'estimation d'impact du chercheur et le besoin d'information du porteur de projets n'est pas issue d'un déplacement de l'intérêt des chercheurs, mais plutôt d'un apprentissage de l'évaluation par les porteurs de projets. C'est parce que les porteurs de projets s'approprient le vocabulaire évaluatif et en comprennent les enjeux qu'ils ont pu, dans plusieurs cas, estimer que les résultats scientifiques ne dépeindraient pas fidèlement le projet et ne satisferaient pas leurs intérêts d'apprentissage. Ces éléments de décalages sont multiples. Ils peuvent porter sur une remise en cause des indicateurs d'impact, ou de la capacité de cette méthodologie à restituer la dimension qualitative du projet et ses dimensions plus globales liées à sa mise en œuvre.

Le premier sujet de désaccord peut survenir au cours de la mise en œuvre du projet quand les porteurs de projets estiment que les indicateurs d'impact, censés estimer l'efficacité du dispositif, ne sont pas pertinents. Cette dimension de désaccord devient problématique particulièrement lorsque les résultats produits par les chercheurs grâce à

⁹⁵⁴ Pour la Mallette des parents par exemple un acteur du rectorat explique : *« son successeur [au recteur] n'était pas du tout sur ces trucs là et on a fait la réunion sans lui »*

ces indicateurs estimaient que le projet n'avait pas un impact significatif. Ces divisions sont apparues dans plusieurs expérimentations, mais le cas de l'expérimentation conduite par la Mission Locale de Tulle est emblématique⁹⁵⁵. Ce projet avait pour objectif la « sécurisation du parcours des jeunes s'engageant dans l'apprentissage ». Cependant, un désaccord est survenu au cours de l'expérimentation autour de ce qui était considéré comme une situation d'échec. Était-ce l'interruption du contrat d'apprentissage - l'indicateur choisi lors de la mise en place du protocole - ou la rupture de parcours, peu importe la formation suivie ? Un des porteurs du projet de la Mission Locale de Tulle explique l'apprentissage qui l'a amené à questionner ces indicateurs : *« Quand on a commencé l'expérimentation, on ne savait pas vraiment mettre des mots sur les choses. Quand je parle de la notion de rupture de parcours et pas de formation, on ne le disait pas comme ça au départ (...). Je parlais d'indicateurs, moi j'ai été amené à construire des indicateurs sur lesquels nous avons beaucoup échangé avec l'équipe d'évaluation, mon indicateur à moi c'est de dire « j'ai un flux de jeunes qui arrivent, quel indicateur va me dire je travaille bien, je travaille mal. Est-ce la rupture brute du contrat d'apprentissage ? Moi je dis rupture de contrat d'apprentissage, ça ne veut pas dire rupture de formation. Ça n'est pas forcément négatif. Il y a des jeunes qui ne retrouvent pas d'entreprise mais continuent la formation et obtiennent leur examen (...). Ceux qui font un Contrat d'apprentissage dans un autre métier, il y a des jeunes, qui quittaient le CFA, pour des congés maternités, est-ce qu'on peut m'imputer ça ? Par exemple sur 21 % qui ont fait l'objet d'un mouvement, 10% sont toujours en formation, [et] la sortie des radars [concerne] moins de 8% du total des apprentis »*⁹⁵⁶.

Un second obstacle à la reconnaissance par les porteurs de projets de la pertinence des résultats produits est l'incapacité de cette méthode à rendre compte du projet dans sa globalité du fait de sa focalisation sur l'estimation de l'impact. Un évaluateur du Crest revient sur ce décalage : *« Je ne suis pas sûr que ce genre de résultats leur parle vraiment tant que ça. Pour nous, ça parle parce qu'on quantifie des hypothèses, on est dans un champ assez quantitatif où il y a des hypothèses théoriques qui portent sur des modèles... ils [les porteurs de projets] sont beaucoup plus sensibles aux études de nature sociologique qu'aux études de nature quantitative (...). C'est vrai que nos résultats sont souvent très modestes, ils sont purement quantitatifs. »*. En effet, plusieurs commanditaires et équipes d'évaluation ont choisi de compléter l'estimation de l'impact menée par la méthode expérimentale par une collecte de données qualitatives permettant de rendre compte du projet, de sa mise en œuvre et de sa réception par les bénéficiaires. Par exemple, dans le cas de l'expérimentation de Pôle Emploi sur les opérateurs privés de placement, les résultats des études qualitatives menées par la Dares et par un cabinet privé ont été davantage utilisées pour éclairer la programmation publique, aussi parce qu'ils étaient disponibles à temps pour la décision. Un interlocuteur de la Dares revient sur le manque d'opérationnalité des résultats produits pour améliorer l'intervention évaluée : *« il faut valoriser ce qui peut être opérationnalisé dans les résultats, il y a une sorte de demande de guide pour savoir qu'est ce que je peux faire de mieux, le qualitatif peut donner un peu de chair. Les chercheurs quantitatifs souvent minimisent un peu, mais c'est vrai que l'aridité du chiffre qui dit ça marche, ça marche pas (rire) pour un décideur il dit « ok, bon, mais encore » »*. Les évaluateurs sont conscients de ce manque de corps dans les résultats produits et du décalage possible entre les résultats qu'ils produisent et le projet en tant que tel. Un évaluateur du Céreq explique : *« je ne pourrais*

⁹⁵⁵ Pour le projet Apipaq, l'évaluateur explique lors d'un entretien qu'il y a eu un désaccord avec un acteur du terrain de l'expérimentation, sur le fait que les jeunes déclarant qu'ils allaient débiter une formation étaient considérés comme des sorties négatives du dispositif du fait qu'ils ne soient pas encore en formation.

⁹⁵⁶ Un évaluateur de ce projet explique en entretien : *« les intérêts s'étaient écartés. La Mission Locale disait que le but était de sécuriser leur parcours par forcément de les conduire à l'apprentissage et là c'est plus dur à évaluer »*

pas vous dire si c'est bien ou pas bien APIPAQ, la seule chose que je peux vous dire c'est que ça a été fait dans l'objectif de faire retourner les jeunes vers une formation qualifiante et que par rapport à cet objectif-là, le contrat n'est pas rempli ».

Cette difficulté à rendre compte du projet de manière fidèle et élargie est d'autant plus grande que les expérimentations organisent l'intervention de nouveaux acteurs voire une nouvelle coopération entre acteurs. Dans ces cas-ci, pour les porteurs de projets, l'apport expérimental est l'élaboration de ce partenariat qui devient souvent pleinement opérationnel au cours de la mise en œuvre du projet. Le porteur du projet de la Mission Locale de Tulle revient sur cette différence d'intérêts : *« On n'évalue pas les mêmes choses, l'évaluateur il est sur une évaluation scientifique, ce qu'on évalue c'est de savoir si ce qu'a fait la Mission Locale à un moment a produit un effet ; ce qu'évaluent les partenaires, c'est si on a bien travaillé, si on est content de travailler ensemble et si ça marche. Mais est-ce que réellement c'est lié à l'activité des Missions Locales ? Ca on n'en sait rien. »*. Cette différence d'objet d'intérêt, entre les parties prenantes, est souvent la source de perceptions différenciées sur l'expérimentation et particulièrement on relève la tendance des porteurs à développer des a priori positifs sur leur expérimentation et être surpris par des résultats négatifs⁹⁵⁷. Cette distanciation apparaît clairement dans la note de restitution finale de l'expérimentation « Tous en Stage » rédigée par l'association « Réussir moi aussi » : *« Les nombreux biais existant dans le cadre de l'évaluation ne nous permettent pas enfin de valoriser le travail réalisé dans le cadre de l'expérimentation, les conclusions quantitatives de l'évaluation étant très éloignées des retours qualitatifs extrêmement positifs et élogieux recueillis auprès des enseignants et chefs d'établissements concernés. »*⁹⁵⁸ Les évaluateurs estiment que ces différences de perceptions constituent des biais pour apprécier l'effectivité du projet et expliquent justement que le dispositif proposé par la méthode expérimentale permet de contrôler ces biais de perception. Un évaluateur du Crest explique : *« la relation était très bonne avec les acteurs, mais les résultats sont très décevants pour eux. On avait un peu d'empathie pour eux. Si j'avais fait une évaluation qualitative, j'en suis bien incapable, je suis sûr que j'aurais fini par trouver quelque chose de positif par ce que je les aime bien, ils sont bons, ils sont motivés, ils sont géniaux, imaginatifs, mais il y a un truc qui ne ment pas : c'est le résultat des enquêtes. »*.

4.2.2. Bilan de l'apprentissage politique issu des résultats produits

L'identification de ces différents facteurs susceptibles d'entraver l'utilisation des résultats par les porteurs de projets appelle un retour sur 13 de nos études de cas afin de rendre compte de l'utilisation des résultats de l'expérimentation par les acteurs de l'action publique⁹⁵⁹. Les porteurs de projets n'ont pas tous les ressources nécessaires pour décider de la pérennisation du projet ou non, la nature du projet et son coût entrent aussi en compte dans ce processus. L'étude de l'ensemble de ces cas montre que toutes les décisions politiques prises à l'issue de l'expérimentation sont allées dans le sens de l'intérêt politique du moment et s'inscrivent dans les pratiques traditionnelles du champ d'action stratégique de l'action publique. Si les résultats de l'expérimentation

⁹⁵⁷ Par exemple on retrouve cet étonnement des porteurs dans le cadre de l'expérimentation « Jeunes diplômés », un évaluateur explique : *« Ils étaient hyper étonnés par ce que les résultats étaient dans certains cas faibles (...) ce n'était pas ceux qu'ils avaient prévu »*

⁹⁵⁸ REUSSIR MOI AUSSI, « Tous en Stage », Note de restitution finale du porteur de projet, Fonds d'expérimentation pour la jeunesse, 2011, p. 10.

⁹⁵⁹ Nous n'avons pas pu avoir l'information sur la continuation du projet AjiLoJe et les résultats de l'expérimentation du Permis de conduire sont sortis trop récemment pour en apprécier l'utilisation

contredisaient l'intérêt des acteurs politiques, ils ont été ignorés ou délégitimés. Ce n'est que lorsqu'ils allaient en partie dans le sens de l'intérêt politique, soit du maintien, soit de la suppression de l'intervention, qu'ils ont été mobilisés comme support à la décision. Nous verrons pour conclure les autres apprentissages des porteurs de projets ou les initiatives d'acteurs de l'action publique de capitalisation des résultats.

4.2.2.1. L'absence de prise en compte des résultats devant les impératifs politiques

Notre étude identifie une absence explicite de prise en compte, ou d'intérêt pour les résultats produits par cette méthode d'évaluation par les porteurs de projets dans la moitié des cas. Les facteurs de cette non-utilisation s'inscrivent dans des logiques politiques et témoignent d'une rupture dans la *relation clinique* liant les intérêts des acteurs de l'action publique et les chercheurs.

D'abord, plusieurs évaluateurs ont été confrontés à des *relations cliniques* perturbées par les mises en agenda de la politique nationale. La décision politique ne pouvant pas attendre la production de résultats scientifiques, la décision de pérenniser ou de généraliser le dispositif s'effectua sans prendre en compte les résultats. On retrouve par exemple cette séparation entre la décision publique et l'expérimentation dans l'expérimentation de Pôle Emploi sur les opérateurs privés de placement. Dès la production de résultats intermédiaires, il est apparu que l'accompagnement renforcé par les opérateurs privés était moins performant que celui proposé par Cap vers l'entreprise (Pôle emploi). Ces résultats ont surpris les commanditaires dont certains n'ont pas reconnu la validité de la démonstration évaluative, et la nécessité de la prise de décision ainsi que le besoin d'appui de Pôle emploi dans ses activités a amené à pérenniser le recours aux opérateurs privés malgré l'absence d'estimation de leur efficacité. Un interlocuteur de Pôle emploi explique : *« les convictions ou les nécessités politiques l'emportent sur les faits, et sur les choses qui pourraient apparaître comme relativement objectives et difficilement contestables (...) On avait de toutes façons tourné la page, dans la mesure où il avait été admis d'emblée, dès la première convention tripartite État-Unedic-Pôle Emploi à sa création que le recours aux organismes privés était maintenu, donc d'une certaine manière il y avait une contradiction radicale peut-être, pourraient se dire certains, entre les conclusions de l'évaluation qui disait « le privé, ça coûte cher (on ne l'avait pas démontré, mais on le savait bien), beaucoup plus cher, et ça rapporte moins. » Et pour autant, on a dit, il faut absolument se donner la possibilité du recours aux opérateurs externes parce que ça va soulager la pression qui s'exerce sur les équipes sur le terrain, et ça personne au fond ne l'a remis en cause »*⁹⁶⁰. Cette décision a d'ailleurs été comprise par les évaluateurs⁹⁶¹.

Plusieurs expérimentations du FEJ sur le thème du décrochage ont été aussi perturbées par la généralisation nationale des plateformes de suivi et d'appui aux décrocheurs très proches des dispositifs innovants mis en place localement afin de mieux repérer ces publics. Dès décembre 2010, il a été décidé de mettre en œuvre le Système interministériel d'échange d'information (SIEI) auquel a succédé la mise en œuvre des

⁹⁶⁰ DARES-POLE EMPLOI, *L'accompagnement renforcé des demandeurs d'emploi : évaluation du recours aux opérateurs privés par Pôle emploi de 2009 à 2011*, 15, coll. Les cahiers études, Direction des études et de l'évaluation, 2013.

⁹⁶¹ Un évaluateur du Crest explique *« on a fait une démonstration intéressante un peu inattendue ça n'a pas empêché que l'Anpe continue dans cette démarche d'OPP à tour de bras avec assez peu de réflexion. Ça ne me gêne pas. La décision des OPP elle est liée à différents enjeux, dont l'impact, et le rapport coût-bénéfice n'est qu'un aspect. On fait face à un afflux croissant de chômage et on ne veut pas embaucher des fonctionnaires, même si on ne sait pas si c'est très efficace à court terme. Après on est dans le champ de la décision politique où il y a tout un champ de considérations, qui ne sont jamais illégitimes, l'évaluation, ne doit pas guider toute la décision publique. »*

plateformes et la Circulaire n°2011-028 du 9 février 2011⁹⁶². Ces expérimentations, dont notamment « APIPAQ », ont dû réorganiser leurs actions autour de ces plateformes. Ces dernières ne correspondaient pas pour autant au dispositif de repérage mis en place par l'expérimentation et il a donc été remplacé par un nouveau système avant la fin de l'expérimentation.

Le cas de l'expérimentation de l'internat d'excellence de Sourdun est aussi intéressant pour illustrer l'importance du portage politique dans la prise en compte des résultats des expérimentations. Cette expérimentation a connu d'abord une généralisation, puis une suppression, après le changement de majorité, avant la production des résultats finaux de l'évaluation. L'internat de Sourdun qui devait tenir lieu de « pilote » des internats a fait l'objet d'une généralisation dès 2010, ce qui amène un rapport de la Cour des comptes à souligner que « *Le processus d'évaluation de cette politique, conçu pourtant dès l'origine, a été éclaté et a eu un impact faible sur le processus de décision* »⁹⁶³. En retour, après le changement de majorité présidentielle, la décision d'arrêter les internats d'excellence a été annoncée le jour où le rapport d'évaluation de l'internat de Sourdun a été publié⁹⁶⁴. Cette décision a été suivie en juin 2013 par l'arrêt des financements du FEJ pour les autres internats d'excellence⁹⁶⁵. Cette initiative constituait en effet une mesure phare en éducation de la présidence de Nicolas Sarkozy, qui a participé à leur inauguration. Elle organisait le retour de l'intervention de l'État central dans la mise en place d'établissements publics nationaux, compétence délocalisée depuis les lois Defferre de 1982⁹⁶⁶. Il s'inscrivait dans l'agenda social du gouvernement montrant que le gouvernement intervenait auprès des jeunes en difficulté, mais « méritant »⁹⁶⁷. Cette mesure, coûteuse et beaucoup trop marquée politiquement a été donc interrompue après le changement de majorité présidentielle. Un évaluateur de ce projet résume cette expérience : « *sur l'internat d'excellence, je comprends bien les différentes dimensions, et la question de l'impact n'est pas la seule qui rentre en ligne de compte* ». Cette non-prise en compte des résultats d'évaluation est aussi présente dans le cas de l'expérimentation de l'accompagnement des RMistes où le porteur du projet a reconduit son contrat avec les opérateurs privés de placement malgré son rapport coût-efficacité jugé négatif par les évaluateurs⁹⁶⁸, de même que pour l'expérimentation « Tous en Stage » où le dispositif continue d'être mis en place malgré l'absence d'identification d'impact par le rapport d'évaluation⁹⁶⁹. Il est en effet, souvent impensable pour les porteurs de revenir en arrière. Comme l'explique un des porteurs de l'expérimentation

⁹⁶² Code de l'éducation - Article L313-8, (2010) Code de l'éducation; DELIBERATION n°2010-448, 2 décembre 2010.

⁹⁶³ « Cette politique, présentée comme expérimentale avec la création de deux lycées d'État, est suivie, dès 2010, d'un plan ambitieux de 20 000 places en internats d'excellence (12 000 places dans des établissements spécifiques créés à cet effet et 8 000 places labellisées dans des internats existants) » COUR DES COMPTES, *Rapport public annuel, Des internats d'excellence à ceux de la réussite: la conduite chaotique d'une politique éducative et sociale*, février 2014.

⁹⁶⁴ En témoignent ces deux articles publiés le 11 avril 2013, l'un dans Le Monde donnant les résultats d'évaluation, l'autre dans le Figaro annonçant la décision de le supprimer Caroline BEYER, « Vers la fin des internats d'excellence », *Le Figaro* (11 avril 2013); Maryline BAUMARD, « L'internat d'excellence fait progresser les élèves en maths, pas en français », *Le Monde*. (11 avril 2013).

⁹⁶⁵ COUR DES COMPTES, *Rapport public annuel, Des internats d'excellence à ceux de la réussite: la conduite chaotique d'une politique éducative et sociale*, février 2014, p. 193.

⁹⁶⁶ En effet comme le souligne le même rapport de la Cour des comptes : « Ces deux premiers internats ont été créés sous le statut d'établissement public national (EPN), cadre juridique exorbitant du droit commun. Depuis les premières lois de décentralisation, en effet, les établissements du second degré sont des établissements publics locaux (EPL) d'enseignement relevant des régions (lycées) et des départements » (collèges) pour leur construction et leur fonctionnement non pédagogique ». COUR DES COMPTES, *Rapport public annuel, Des internats d'excellence à ceux de la réussite: la conduite chaotique d'une politique éducative et sociale*, février 2014, p. 181.

⁹⁶⁷ On lit dans un article du Figaro une partie du discours inaugural de Nicolas Sarkozy : « Ces structures s'adressent à des jeunes méritants, pas forcément des premiers de classe, mais qui se battent malgré des difficultés immenses » Marie-Estelle PECH, « Nicolas Sarkozy inaugure un «internat d'excellence» », *Le Figaro* (9 septembre 2010).

⁹⁶⁸ On peut voir sur le site internet du prestataire que le contrat a été reconduit de novembre 2008 à mars 2010 <http://www.ingeus.fr/nos-accompagnements-sur-mesure/allocataires-rsa/diagnostic-employabilite-orientation/>

⁹⁶⁹ On lit sur le site internet de l'association pour l'année succédant aux résultats : « Année 2013 / 2014 : Continuité du dispositif dans les départements 75, 93 et 94. Redémarrage du dispositif dans 5 collèges du 91 ». <http://reussirmoiaussi.fr/references.htm>

de Tulle, une expérimentation n'est pas une « parenthèse », mais une « étape ». Il explique : « *Le travail s'est stabilisé maintenant il rentrerait dans l'esprit de personne de modifier les pratiques (...). L'expérimentation a apporté une culture commune* ». L'association Actenses qui a bénéficié du FEJ comme financement pour lancer son action semble aussi continuer ses interventions de Parrainage grâce au financement d'entreprises privées, malgré les réserves des évaluateurs sur l'effectivité de l'expérimentation⁹⁷⁰.

À l'étude de ces différents cas, il semble donc que l'expérimentation aléatoire ait été mobilisée dans une préoccupation tactique de légitimation de l'intervention, mais l'intervention a été maintenue sans rapport avec ses résultats, mais plutôt suivant l'intérêt politique, les possibilités de financement, ou des phénomènes de dépendance au sentier. Lorsque les résultats scientifiques tardent à être produits ou qu'ils n'étaient pas le point de vue des commanditaires ils sont mis de côté par les acteurs de l'action publique. Ainsi dans les cas cités, aucun résultat contradictoire avec l'agenda politique n'a contribué à le modifier. Le maintien des dispositifs expérimentés peut être aussi issu des phénomènes de dépendance au sentier (*path dependency*) vis-à-vis des *process* mis en place lors de cette expérimentation. Pour plusieurs porteurs, le financement expérimental a permis de lancer l'action et le facteur le plus déterminant de sa pérennisation était l'identification de nouvelles sources de financement pour ce projet.

4.2.2.2. La prise en compte des résultats soutenant la décision politique

Passons maintenant à l'étude des cas de décisions publiques allant dans le sens des résultats de l'évaluation. On peut observer cette prise de décision après l'expérimentation seulement dans deux cas et auprès des porteurs de projets qui étaient en position d'être décisionnaires sur cette pérennisation.

On retrouve cette utilisation vertueuse des résultats d'évaluation pour l'expérimentation AQ3E, dont les résultats ont été positifs puisque le dispositif augmente la qualité des emplois sans pénaliser les résultats scolaires⁹⁷¹. Pour ce projet, le porteur, ici l'Université du Maine, pouvait lui-même prendre la décision de pérenniser le dispositif qui d'ailleurs s'auto-finançait en partie grâce aux rentrées de taxes d'apprentissages qui accompagnaient ce partenariat avec les entreprises⁹⁷².

La « mallette des parents » est présentée comme le modèle vertueux d'expérimentation ayant donné lieu à une généralisation à 1300 collèges en 2010, au vu de ses résultats positifs⁹⁷³. Néanmoins, la généralisation de cette expérimentation s'inscrit dans le contexte particulier de l'accession du porteur du projet, le recteur de Créteil, au poste de DGESCO (directeur général de l'enseignement scolaire) au ministère de l'Éducation Nationale en décembre 2009. Celui-ci précise : « *Comme je suis devenu DGESCO dans l'intervalle, on s'est dit qu'il fallait le développer dans toute la France, d'autant plus qu'on voyait toutes les vertus, l'une d'entre elles étant que ce n'était pas très coûteux par rapport à l'effet choisi. Donc on attendait l'évaluation pour envisager cette*

⁹⁷⁰ Nous avons tenté à plusieurs reprises sans succès d'avoir un entretien avec ces porteurs de projet, et nous ne sommes donc pas en mesure d'apporter davantage de précision à ce sujet.

⁹⁷¹ Les évaluateurs conseillent sa généralisation : « *Cette expérimentation gagnerait à être généralisée à d'autres établissements d'enseignement supérieur qui sont concernés par la problématique du travail salarié des étudiants* ». LABORATOIRE ERUDITE, *Améliorer la Qualité des Emplois Exercés par les Étudiants (AQ3E)*, Rapport d'évaluation, 2013, p. 25.

⁹⁷² L'évaluateur de ce projet explique en entretien « le président de l'université a décidé de pérenniser le dispositif, L'expérimentation s'arrête le 30 juin, mais il va perdurer avec le nouveau poste créé »

⁹⁷³ Mathieu VALDENNAIRE et Jean BERARD, « L'expérimentation pour renouveler les politiques publiques ? », *La Vie des idées* (25 juin 2013).

généralisation que j'ai proposée au ministre, qui a accepté. (...) Le fait que moi-même je sois devenu une sorte d'abeille en passant de l'échelle Créteil à l'échelle nationale est un cas particulier ». Cependant, cette généralisation s'inscrit dans le contexte de la loi Ciotti passée en septembre 2010 qui plaçait la participation des parents à la scolarité de leurs enfants en haut de l'agenda politique⁹⁷⁴. Un des membres de l'équipe des porteurs de projets explique : « *Martin Hirsch était clairement contre la loi Ciotti et pour la Mallette des parents, et Nicolas Sarkozy l'a repris à son compte pour dire : « on fait la loi Ciotti, mais on fait aussi la Mallette des parents, ce n'est pas comme si on était uniquement sur la punition* ». Cette généralisation a été fortement soutenue par le FEJ, même financièrement, afin de l'ériger en exemple de bonne pratique expérimentale⁹⁷⁵.

Il semble ainsi que la généralisation de cette expérimentation tient davantage d'une logique politique plutôt que d'une pratique scientifique où une généralisation s'effectue au regard de la validité externe des résultats produits. En effet, en regardant de plus près cette expérimentation, il apparaît que la durée d'expérimentation a été réduite de 3 ans à 2 ans et que l'intervention de la « Mallette » en tant que telle n'a été évaluée rigoureusement que lors de la première année d'intervention sur un échantillon de 200 classes dont 100 tests, parmi lesquelles 543 parents déjà volontaires ont été sollicités et dont au final 58% ont participé à au moins une réunion sur trois. En effet, même si l'expérimentation de la deuxième année semble conforter les résultats très prometteurs sur les jeunes et les parents de la première année, l'objet de cette expérimentation reste sur les modes de mobilisation de ces derniers et ne mesure donc pas précisément l'impact des réunions « Mallette des parents »⁹⁷⁶.

Aussi, remarquons que dans ces deux cas d'utilisation des résultats de l'évaluation, les deux porteurs de projets appartiennent au groupe des *titulaires* du champ stratégique de l'action publique identifié précédemment, et qu'ils disposent donc de la légitimité ainsi que de la marge décisionnelle et potentiellement financière suffisante pour décider soit de la pérennisation du projet soit de sa généralisation. De plus, ces deux projets sont peu coûteux et ne perturbent pas fortement les pratiques des acteurs de terrain. La forte participation des évaluateurs à la conception de ces deux projets, dont un a été exclusivement conçu par l'équipe d'évaluation (AQ3E), a permis une structuration du projet favorisant sa généralisation.

Cette utilisation des résultats est beaucoup plus complexe lorsque les porteurs de projets appartiennent aux deux autres groupes du champ d'action stratégique de l'action publique. Les missions locales (du groupe des *titulaires précaires*) ou les associations (groupe des *challengeurs*), ont besoin de mobiliser de nouveaux financements afin de pérenniser leur projet, et nous avons vu précédemment que cette démarche était fragilisée par la disponibilité tardive des résultats d'évaluation. Le défi pour ces structures est alors d'arriver à attirer l'attention de potentiels financeurs sur leur expérimentation afin qu'ils en financent la généralisation ou pérennisation.

⁹⁷⁴ Pour mémoire la loi Ciotti, proposait de lutter contre l'absentéisme scolaire en supprimant les allocations familiales aux parents des enfants ayant au moins 4 demi-journées d'absence irrégulière à l'école dans le mois. Elle fut abrogée en janvier 2013.

⁹⁷⁵ Le même interlocuteur souligne : « *Le FEJ aussi poussait très fort sur la généralisation pour pouvoir montrer que les expérimentations ce n'était pas que « ça s'arrête et on passe à autre chose ». Mais que c'était pour guider l'action publique et donner des directives. La première année ils nous ont financé une partie de la généralisation ce qui était à la limite de leurs compétences quand même, mais justement pour pouvoir montrer le cycle de l'expérimentation* »

⁹⁷⁶ Le tirage au sort porte sur le mode de sollicitation, et on ne peut plus comparer l'impact du programme entre des classes test ou témoin, mais qu'entre des classes ayant un différentiel de participation des parents de 5 points, ce qui porte les évaluateurs à être beaucoup plus réservés sur ces résultats, qu'ils estiment non exempts d'aléas statistiques potentiels ECOLE D'ECONOMIE DE PARIS, *Quels effets attendre d'une politique d'implication des parents dans les collèges, évaluation de l'impact de la Mallette des parents*, Fonds d'expérimentations pour la jeunesse, 2011.

Ainsi, un rapport national avait envisagé la généralisation du dispositif d'accompagnement des décrocheurs proposé par l'expérimentation « Rebond » porté par l'AFIJ, sous réserve de l'effectivité de son impact⁹⁷⁷. Cet impact est au final mitigé, mais estimé comme positif sur les actions d'accompagnement alliant contractualisation avec les jeunes décrocheurs et orientation vers les métiers en tension. Néanmoins, comme nous l'avons évoqué cette structure n'a pu survivre à la période d'incertitude d'apports de subventions d'État et a dû fermer en décembre 2013.

La Mission Locale de Beaune avait aussi amorcé une entreprise de mobilisation des potentiels financeurs afin de pérenniser son projet voire de le généraliser à d'autres Missions Locales, puisque son impact avait été estimé comme probant sur le secteur de l'hôtellerie/ restauration. Le porteur du projet, avait reçu Bertrand Martinot le DGEFP à l'époque, et Jean Philippe Daubresse qui succéda à Martin Hirsch au ministère de la jeunesse et des solidarités actives, pour leurs présenter le projet. Il était intervenu au conseil d'orientation pour l'emploi sur la problématique de l'emploi des jeunes en janvier 2011⁹⁷⁸. La réflexion avait été amorcée sur une extension progressive de ce programme en Bourgogne, la Mission Locale de Beaune avait été chargée de conduire une réflexion pour le Comité régional à l'insertion des Jeunes (CORIJ) et l'accord de principe avait été réalisé le 6 janvier 2012⁹⁷⁹. Or, lorsque nous avons tenté de contacter le porteur du projet en avril 2013 la Mission Locale était menacée de fermeture faute d'une baisse de 25% de sa dotation d'État⁹⁸⁰. Elle reste encore en activité grâce à une aide exceptionnelle de la Région et de l'État, mais sa situation reste précaire⁹⁸¹.

Deux expérimentations ayant abouti à un impact non probant de l'intervention, voire à un effet négatif ont mené à l'abandon du projet. Néanmoins, pour ces deux cas, la démonstration de l'inefficacité du projet allait dans le sens soit de l'intérêt politique qui ne tenait pas à mettre en œuvre cette intervention, soit de la perte d'intérêt politique pour l'intervention. Le premier exemple est celui de l'expérimentation de l'accompagnement des « jeunes diplômés » dont l'efficacité globale sur l'insertion des jeunes n'a pas été démontrée, et qui comme nous l'avons vu précédemment, suite à un changement de cabinet ministériel a perdu l'attention des instances décisionnaires. Le second exemple est celui du CV anonyme. Cette expérimentation, issue de la Loi pour l'Égalité des chances de 2006 qui stipulait que toutes les entreprises de moins de 50

⁹⁷⁷ On lit dans la proposition 11 du rapport contre le décrochage scolaire : « *Le recteur pourrait assurer en maîtrise d'œuvre directe la coordination de l'offre de réorientation au sein de l'espace académique d'enseignement supérieur et déléguer l'opérationnel de la mission de réorientation vers l'emploi à l'AFIJ sur la base de l'expérience menée par cette dernière dans le cadre des projets Hirsch. Sous réserve d'une évaluation positive, le travail effectué par l'AFIJ auprès des étudiants décrocheurs de plusieurs universités pourrait être reproduit sur l'ensemble du territoire national.* » Christian DEMUINCK, *Le décrochage universitaire*, Ministère de l'Enseignement supérieur et de la Recherche, 2011, p. 106.

⁹⁷⁸ MISSION LOCALE DE BEAUNE, « *Inscrire les contrats en alternance dans une logique de parcours sécurisé* », Note de restitution finale, Fonds d'expérimentation pour la jeunesse, 2012.

⁹⁷⁹ On peut retrouver la présentation de cette intervention sur le site du COE : http://www.coe.gouv.fr/IMG/pdf/R._Barbier_-_2011-01-06.pdf, les autres modalités d'organisation de la communication autour de l'expérimentation sont mentionnées dans le document MISSION LOCALE DE BEAUNE, « *Inscrire les contrats en alternance dans une logique de parcours sécurisé* », Note de restitution finale, Fonds d'expérimentation pour la jeunesse, 2012, p. 16.

⁹⁸⁰ On lit sur le site de la gazette des communes : « *La Direction régionale des entreprises, de la concurrence, de la consommation, du travail et de l'emploi (Direccte) de Bourgogne a annoncé en mars dernier la réduction, à compter du 1er janvier 2013, de la dotation de fonctionnement. (...) La baisse s'élevait en mars à plus de 100 000 euros (pour atteindre - 25,8 % sur trois ans). Les discussions avec l'État, fin mai, l'ont ramenée à 27 720 euros, avec une compensation du Fonds paritaire de sécurisation des parcours professionnels (FPSPP) assujettie à des objectifs (28 contrats ANI d'ici à fin juin).* » <http://www.lagazettedescommunes.com/170269/la-mission-locale-de-beaune-fragilisee-par-une-redistribution-des-moyens-par-la-direccte/>

⁹⁸¹ Un journal de Saône et Loire, *Le bien public*, revient en janvier 2014 sur la situation de la Mission Locale <http://www.bienpublic.com/edition-de-beaune/2014/01/13/un-equilibre-encore-fragile>

salariés devaient avoir recours au CV anonyme pour recruter⁹⁸². Cependant l'administration n'avait pas fait passer le décret d'application. Le gouvernement avait donc décidé de demander à Pôle Emploi de mettre en place un protocole d'évaluation afin de savoir si le CV anonyme permettait de lutter contre les discriminations⁹⁸³. Pour autant, ce dispositif n'était ni porté par la présidence qui avait jusqu'à présent retardé son application, ni par Pôle Emploi dont un des interlocuteurs rencontrés explique : « *on avait la conviction que généraliser le CV anonyme était une bêtise, une erreur de politique publique* ». Les résultats de l'évaluation ont été mitigés et difficiles à interpréter pour les chercheurs. Ils établissaient l'efficacité du CV anonyme pour lutter contre l'homophilie lors du recrutement, mais estimaient un effet négatif pour réduire les discriminations envers les candidats d'origine étrangère ou issus de l'immigration dans le cadre des entreprises ayant participé à l'expérimentation. Cette complexité liée à l'interprétation des résultats a contribué à brouiller leur communication et conduit à leur instrumentalisation par les acteurs publics échappant au contrôle des évaluateurs. Un évaluateur de cette expérimentation explique : « *Une des lectures possibles qu'on avait vraiment exclues noir sur blanc dans le rapport dans son introduction et pas aussi clairement dans sa note de synthèse, c'est la lecture qui disait que ce rapport en réalité disait que les entreprises françaises ne discriminent pas envers les personnes de nationalités étrangères ou issues de l'immigration, donc là c'est intéressant d'aller voir la chronique d'Éric Zemmour à ce sujet qui a participé au relais qui a dit « finalement les entreprises françaises ne sont pas discriminantes et à la limite ces pauvres Français nés en France de parents français sont discriminés* »⁹⁸⁴. *Ca c'est une interprétation impossible puisque le test de ça, c'est par du testing en aveugle, et c'est pas ça qu'on avait fait*⁹⁸⁵. (...) *Je pense que la lecture raisonnable de nos résultats c'est de dire : c'est un dispositif effectivement intéressant pour les discriminations contre le genre, et contre les discriminations selon l'origine il est difficile à dire, mais en tous cas, pour certaines entreprises ça semble plutôt contre-productif* ». En effet, il semble à la lecture de ce rapport et de nos entretiens que ce soit, davantage les modalités de l'anonymisation du CV qui ne soient pas pertinentes dans la lutte contre les discriminations puisqu'elles privent les recruteurs de l'information pour interpréter certaines faiblesses dans le CV.

Ainsi ces deux exemples de décision d'arrêt d'un dispositif suite à leur non-efficacité ne témoignent pas d'un processus de persuasion de la décision publique par la preuve, mais plutôt de résultats parfois instrumentalisés, allant dans le sens de l'intérêt politique contre l'expérimentation, ou de son désintérêt pour cette dernière.

⁹⁸² LOI n° 2006-396 pour l'égalité des chances, (2006) 2006-396.

⁹⁸³ Source entretien avec un interlocuteur de Pôle Emploi : « *Donc, le gouvernement et notamment Azouz Begag ont dit « eh bien, on ne prendra de décrets que sur la base de résultats d'une évaluation qui nous permet de savoir si vraiment le CV anonyme a les effets tels qu'on peut les attendre, et cette évaluation on la confie à Pôle Emploi* ».

⁹⁸⁴ Les propos exacts de ce dernier peuvent être toujours entendus sur le site d'RTL, quelques morceaux choisis : « *Ce travail dément avec une rare cruauté la thèse imposée et enseignée sans relâche depuis des années dans les écoles comme dans les médias la fameuse discrimination (...) Pied de nez formidable à cette accusation lancinante* » « *comme le dit le responsable de l'étude* » « *certain pourrait les [les recruteurs] accuser de discrimination positive, de rupture de l'égalité au détriment des Français de souche* ». Cette étude « *Prouve que la France a mis en place la plus grande machine à intégrer les étrangers* », « *Les résultats de cette étude vont conforter le gouvernement auquel on reprochait de n'avoir jamais pris le décret d'application de cette mesure qui figurait dans la loi sur l'égalité des chances votée en 2006 après les émeutes de l'automne 2005* ». « *Eric Zemmour : "Les CV anonymes reposent sur des présupposés d'une France raciste"* », RTL. (13 avril 2011).

⁹⁸⁵ En effets d'autres expérimentations conduites par des méthodes de *testing*, c'est-à-dire d'envoi de faux CV où seule la variable de l'origine varie permettent de mettre en valeur cette discrimination à l'embauche Pascale PETIT, Emmanuel DUGUET et Yannick L'HORTY, *Discrimination résidentielle et origine ethnique: Une étude expérimentale en Ile-de-France*, TEPP Rapport de recherche, octobre 2012.

4.2.2.3. Les apprentissages en cours

Plusieurs apprentissages issus des expérimentations par assignation aléatoire sont encore en cours et il est difficile dans ce travail d'en apprécier pour l'instant rigoureusement les effets.

Le fonds d'expérimentation qui, comme nous l'avons vu précédemment, promouvait une utilisation de type *conceptuelle* des résultats d'expérimentation - c'est-à-dire cumulative et faisant la synthèse de plusieurs interventions pour éclairer la décision publique - a organisé plusieurs réunions thématiques ainsi que des dossiers de synthèse afin de promouvoir l'utilisation des résultats. Un interlocuteur du FEJ responsable de la capitalisation explique : « *l'idée de ces ateliers, c'est de commencer à voir sur quoi on allait pouvoir capitaliser, on y invitait aussi les parties prenantes du ministère thématique pour qu'ils s'y intéressent* ». Le plan de capitalisation élaboré par la Mafej sous consultation du Conseil Scientifique consiste principalement à rendre publics les résultats d'évaluation et à les présenter sous forme de synthèses thématiques. De décembre 2010 à juin 2011 des cahiers thématiques regroupant les expérimentations par axe, tel que le « décrochage scolaire » par exemple, ont été constitués et publiés sur le site internet du FEJ afin de répertorier les différentes expérimentations et leurs modes d'évaluation. Néanmoins, à ce stade, aucun résultat n'était encore disponible⁹⁸⁶. En octobre 2013, un document de synthèse des résultats d'expérimentation a été publié et revient de manière thématique et transversale sur les premiers enseignements issus des expérimentations⁹⁸⁷. Cependant, le temps de production des résultats d'expérimentation, et celui de la constitution de ces synthèses les inscrivent dans une période trop contemporaine pour en apprécier l'effectivité.

Le FEJ publie sur son site internet, la quasi-totalité des rapports d'évaluation d'expérimentation permettant à chacun de s'en saisir. Comme l'explique le responsable de la capitalisation à la Mafej, « *les hommes politiques font ce qu'ils veulent, donc nous, dans le plan de capitalisation on a essayé d'avoir une approche un peu large des destinataires (...). Après, se saisissent des rapports les gens qui veulent s'en saisir* ».

Cette communication autour des résultats est aussi organisée par les équipes d'évaluation, autant par le J-PAL que par le Credoc et le Céreq. L'une des missions du J-PAL, dont la devise est « *translating research into action* » (traduire la recherche en action), est en effet de proposer des supports de communication adaptés aux décideurs publics et de favoriser la diffusion des résultats. Il répertorie sur son site, les différentes expérimentations menées par ses affiliés et produit parfois des plaquettes de communication (appelés *Briefcase*) destinées aux décideurs publics comme pour le cas de la Malette des parents en février 2011⁹⁸⁸. Comme nous l'avons précisé plus haut, le Credoc et le Céreq ont organisé soit des conférences de restitutions, ou dédié des articles dans leurs revues institutionnelles aux expérimentations qu'ils ont conduites, permettant la diffusion à d'autres acteurs publics. Par ailleurs, des porteurs de projets qui ont mené

⁹⁸⁶ Les thématiques sont les suivantes : « sécuriser l'orientation vers l'alternance, prévenir les ruptures » (avril 2011), « prévention et lutte contre le décrochage scolaire » (mars 2011), « soutenir et valoriser l'engagement des Jeunes » (mars 2011), « développement du microcrédit pour les jeunes et soutien à l'entrepreneuriat » (mars 2011), « Améliorer l'accès aux soins des jeunes » (février 2011), « orientation » (décembre 2010), « jeunes suivis par l'aide sociale à l'enfance et la protection judiciaire de la jeunesse, et des jeunes sous main de justice » (juin 2011), « livret expérimental de compétences » (mai 2011). Ces livrets thématiques sont disponibles sur le site du FEJ : <http://www.experimentation.jeunes.gouv.fr/1146-dossiers-de-presentation-des-experimentations.html>

⁹⁸⁷ FONDS D'EXPERIMENTATION POUR LA JEUNESSE, *Premiers enseignements des expérimentations, synthèses thématiques 2013*, Ministère des Sports, de la Jeunesse, de l'Education populaire et de la Vie associative, 2013.

⁹⁸⁸ Document consultable sur le site du J-PAL Europe : <https://www.povertyactionlab.org/fr/publication/getting-parents-involved> ABDUL LATIF JAMEEL POVERTY ACTION LAB, *Getting parents involved*, J-PAL Briefcase, février 2011.

plusieurs expérimentations ont parfois publié leurs analyses, comme l'Association Nationale des Directeurs de Mission Locale (ANDML) par exemple qui a produit un document de synthèse proposant une analyse croisée des connaissances acquises sur l'efficacité de leur intervention et les modalités de mise en œuvre⁹⁸⁹.

La capacité ou volonté de la puissance publique à se saisir de ces documents n'est pas encore avérée. Le gouvernement issu du changement de majorité présidentielle a posé la politique de jeunesse comme une priorité. Le Premier ministre Jean-Marc Ayrault a confié à Valérie Fourneyron, ministre chargée de la Jeunesse, l'organisation d'un Comité interministériel de la jeunesse en précisant que « *la mise en œuvre d'une politique publique de la jeunesse efficace et adaptée aux exigences sociales et économiques constituait une priorité forte du gouvernement.* »⁹⁹⁰. Ce comité s'est réuni le 21 février 2013, il a fixé la feuille de route de la politique de la jeunesse dans un rapport « *priorité jeunesse* ». On pourrait faire l'hypothèse que la mission « d'efficacité » de cette politique pourrait s'appuyer en partie sur les travaux du Fonds d'expérimentation pour la Jeunesse, il ressort que ce document ne fait qu'une fois explicitement référence à ce Fonds⁹⁹¹. Il ne mentionne que les résultats des travaux d'expérimentations des Missions Locales sur les jeunes sous main de justice et le livret de compétence expérimental (autre expérimentation du FEJ) est évoqué sans référence à son expérimentation⁹⁹². Plusieurs nouvelles expérimentations sont annoncées et organisées hors du cadre du FEJ, comme l'expérimentation de la participation des parents à l'orientation en fin de 3^{ème}, l'intermédiation locative pour les jeunes, les « emplois francs » et l'accompagnement des jeunes en apprentissage.

Enfin, d'autres apprentissages plus subtils sont issus de la participation des porteurs de projets aux expérimentations, ils ne sont pas particuliers à la méthode par assignation aléatoire, mais méritent néanmoins d'être mentionnés. Ces apprentissages tiennent à la sensibilisation des acteurs de l'action publique à la culture évaluative, et notamment l'intérêt de partenariat avec les évaluateurs externes mobilisant des méthodes scientifiques. Ainsi le porteur du projet de l'expérimentation APIPAQ explique : « *je pense qu'il faut le renouveler, mais pas forcément avec la même méthode. Ce qui est intéressant c'est quand un extérieur évalue notre action pour avoir un autre regard, moi la méthode aléatoire me plaît bien, mais il faut qu'elle soit invisible. On est habitué à l'auto-évaluation, on s'évalue nous même sur la base de tableaux de bord* ». Certains ont aussi compris l'intérêt de participer plus activement au processus de choix méthodologiques et de définition d'indicateurs, afin que ces derniers témoignent davantage de leur action sur le terrain. Ils ont dans l'ensemble apprécié ce moment de prise de recul sur l'action leur permettant de redéfinir leurs objectifs. Un porteur du projet de la Mission Locale de Tulle explique : « *la prochaine fois, je serai plus vigilant sur la méthodologie employée par l'expérimentateur, car je ne vais pas totalement lui déléguer le truc dans la mesure où il ne connaît pas mon métier. Quand ça ne me semble pas praticable, même si lui pense que ça l'est, il faut que je défende mon bout de gras. (...) J'ai acquis du recul sur les questions d'évaluation, de partenariat et de conduite des politiques publiques. Avec des évaluateurs de ce niveau-là, on prend du plaisir.* ». On peut donc émettre l'hypothèse que la conduite d'expérimentation et leur évaluation

⁹⁸⁹ ASSOCIATION NATIONALE DES DIRECTEURS DE MISSION LOCALE, *Quels enseignements retenir des expérimentations sociales sur l'alternance menées par les Missions Locales?*, 2013.

⁹⁹⁰ COMITE INTERMINISTERIEL DE LA JEUNESSE, *Priorité Jeunesse*, 2013, p. 77.

⁹⁹¹ « *Les expérimentations menées dans le cadre du Fonds d'expérimentation pour la jeunesse (FEJ) ont montré que les freins étaient multiples (coût, maîtrise d'une langue étrangère, influence de l'entourage, niveau scolaire) et que la conjugaison de plusieurs d'entre eux rend souvent la mobilité improbable* » COMITE INTERMINISTERIEL DE LA JEUNESSE, *Priorité Jeunesse*, 2013, p. 46.

⁹⁹² COMITE INTERMINISTERIEL DE LA JEUNESSE, *Priorité Jeunesse*, 2013, p. 40.

ont permis une acculturation progressive des acteurs de l'action publique à la pratique évaluative. Au final, pour tous ces projets, l'évaluation était entièrement financée par le FEJ, et devant les difficultés financières de plusieurs porteurs de projets, il semble pour l'instant difficilement envisageable qu'ils prennent cette dépense à leur charge.

Ainsi, à partir de l'étude de ces différents cas, on observe à ce jour une très faible utilisation par les porteurs de projets des résultats d'évaluations expérimentales par assignation aléatoire. Du fait des difficultés liées à la mise en œuvre des expérimentations, les résultats « tranchés » sur l'efficacité des interventions ont été rarement produits. Ouvrant ainsi parfois la porte à une interprétation politique non maîtrisée par les évaluateurs. Cette utilisation a été pour l'instant restreinte à des cas où les résultats de l'expérimentation correspondaient à un agenda politique et où le porteur de projets avait les compétences de décider ainsi que de financer la pérennisation ou la généralisation du projet. Les estimations de l'impact des dispositifs produits par la méthode par assignation aléatoire sont donc utilisées principalement pour justifier une décision politique déjà prise au préalable et dans le cas d'une volonté de pérenniser l'intervention. L'organisation des expérimentations a pour principale fonction de stabiliser et d'ajuster la mise en œuvre de cette dernière. Ce type d'utilisation sort cet instrument de son utilisation vertueuse et loin d'importer des pratiques scientifiques dans le champ d'action stratégique de l'action publique, il semble davantage relever d'une plus grande dépendance des pratiques scientifiques aux logiques de ce dernier.

Néanmoins, à ce jour, ces résultats restent provisoires du fait de la contemporanéité de notre objet d'étude et de la stratégie cumulative de moyens à long terme défendue par les utilisateurs de cette méthodologie.

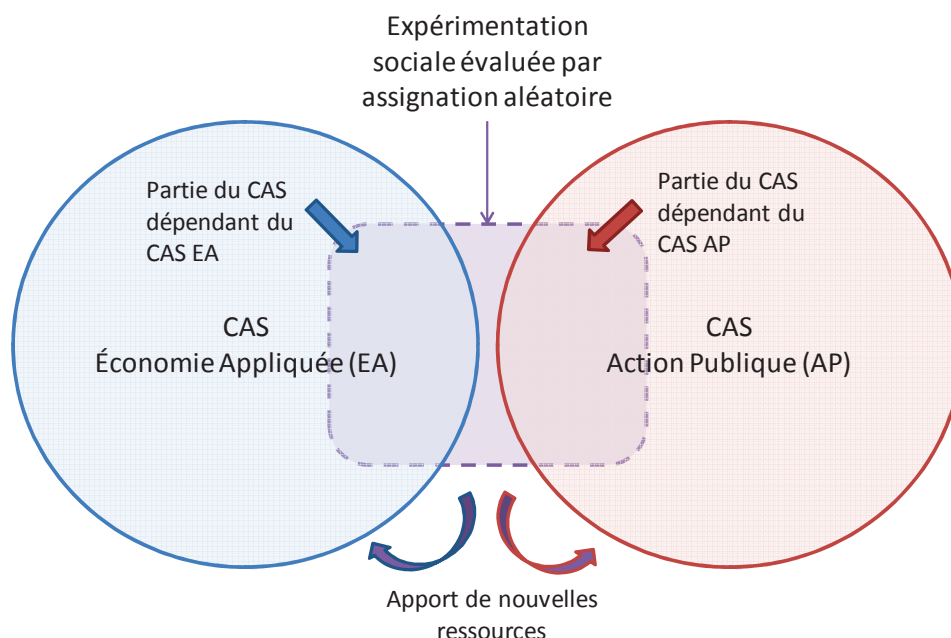
Conclusion du chapitre 3

Ce dernier chapitre étudie la *relation clinique* organisée par l'expérimentation sociale évaluée par la méthode expérimentale par assignation aléatoire. Deux interrogations guidaient cette recherche empirique sur une quinzaine de cas d'expérimentations aléatoires. La première portait sur l'identification de nouveaux modes d'interaction entre les chercheurs en économie appliquée et les acteurs de l'action publique par le biais de cette pratique. La seconde s'intéressait à l'adaptabilité de cette méthode d'évaluation aux modes d'apprentissage des acteurs de l'action publique. Cette conclusion propose de revenir sur nos hypothèses de recherche ainsi que sur la capacité de notre appareillage théorique à les appréhender.

Notre première préoccupation était de comprendre et d'expliquer les motivations de l'engagement des parties prenantes de l'expérimentation aléatoire dans cette entreprise. L'étude de cet objet permet de dégager les intérêts de chacun dans la conduite de cet instrument et notamment d'identifier l'équilibre des variables liées au *Puzzling* (recherche de résolution de problème) et *Powering* (recherche de pouvoir et de captation de ressources) pour ces acteurs. Le concept de *champ d'action stratégique*, qui guide l'attention du chercheur vers l'identification de champs d'appartenances des différents acteurs influençant leur stratégies et intérêts, s'est avéré utile pour comprendre la diversité de ces motivations, élaborer des typologies d'acteurs et expliquer par la suite leur comportement. Ce cadre conceptuel permet notamment d'identifier des variables explicatives communes des comportements des différents acteurs au-delà de leurs appartenances institutionnelles et de rendre compte de l'articulation de leurs intérêts selon la position qu'ils occupent dans le champ d'action stratégique dans lequel ils s'inscrivent. On ne peut comprendre les enjeux au sein de l'expérimentation aléatoire, si on ne s'intéresse pas à son macro-environnement. Elle organise la rencontre du champ

d'action stratégique de l'économie appliquée et de l'action publique et s'inscrit dans une relation de dépendance vis-à-vis des dynamiques de ces derniers. En retour, elle constitue un lieu de nouvelles ressources pour améliorer la position de ses membres dans leur champ d'appartenance respectif.

Figure 17: Les champs d'action stratégiques (CAS) influençant le CAS de l'expérimentation sociale par assignation aléatoire



Lecture : Ce graphique représente le champ d'action stratégique de l'économie appliquée (CAS EA) en bleu et le champ d'action stratégique de l'action publique (CAS AP) en rouge. Bien sûr, cette représentation est simplificatrice puisque nous estimons que chacun de ces champs comprend aussi plusieurs autres champs d'action stratégiques en son sein. Ces deux champs se rencontrent par la conduite d'expérimentation sociale évaluée par la méthode expérimentale par assignation aléatoire et l'influencent par leurs dynamiques de champs. En retour, celle-ci constitue un apport de nouvelles ressources pour ces derniers.

Cette étude empirique appelle néanmoins, à affiner les typologies d'acteurs proposées par les auteurs de ce concept afin qu'elles deviennent davantage opérationnelles. L'opposition *Titulaires/Challengeurs* est apparue trop binaire pour rendre compte des positions dans le champ et nous identifions une position intermédiaire de titulaires moins compétents (au sens de *less skilled*) dans le champ mais bénéficiant tout de même d'un accès privilégié aux ressources. Cette plus grande fragilité est liée soit à la renégociation périodique de leur position par des logiques contractuelles (comme c'est le cas pour la catégorie des *titulaires précaires* du champ de l'action publique), soit à une dépendance envers un groupe de titulaire plus compétent qui leur permet l'accès aux ressources et à une position privilégiée (comme c'est le cas de la catégorie des *titulaires satellitaires* du champ de l'économie appliquée). Ils n'ont pas une position de challengeurs car ils partagent une communauté d'intérêts avec l'ensemble *titulaires*.

Cette confrontation empirique valide notre hypothèse stipulant que cette méthode d'évaluation agit comme une institution sociale organisant de nouveaux modes d'interaction entre les acteurs de l'économie appliquée et ceux de l'action publique. L'étude de ces incarnations diverses sur les différents *sites institutionnels*, conforte la conception de l'institution de T. Lawrence et R. Suddaby comme une structure

malléable et transformée au gré du *travail institutionnel* des acteurs la constituant. Dès lors, des constantes apparaissent dans le type d'interactions organisé par cette dernière.

D'abord, cette méthode augmente la dépendance des chercheurs aux acteurs de l'action publique les amenant à davantage s'investir dans cette sphère. A la différence d'autres méthodes d'économie appliquée, cette méthode scientifique requiert l'accord et le soutien des acteurs de l'action publique pour être conduite par des chercheurs en économie appliquée. Cette méthode implique aussi une plus grande proximité de la recherche en économie appliquée à l'égard du terrain de l'action publique par la nécessité de collecte de données primaires.

En retour, cette institution tend vers l'augmentation de l'ingérence des chercheurs dans la conduite de l'action publique. Pratiquée essentiellement par des chercheurs, elle vient solliciter l'intérêt académique de ces derniers pour les interventions qu'ils évaluent. Les évaluateurs choisissent d'évaluer des projets pouvant nourrir les débats d'économie appliquée d'actualité. Ils modifient systématiquement à la marge des éléments du traitement pour qu'ils correspondent à leur objet d'intérêt voire éventuellement initier les projets expérimentaux. La pratique de l'assignation aléatoire nécessite une manipulation expérimentale des bénéficiaires et de la mise en œuvre en tentant de neutraliser la subjectivité des acteurs de terrain. Bien que des évaluateurs d'expérimentations sociales mobilisant d'autres méthodes ont aussi signalé leur intervention auprès du porteur de projets, celle-ci était perçue comme un dysfonctionnement alors que pour la méthode aléatoire, elle constitue un pré-requis. De fait, on observe une *relation clinique* déséquilibrée entre les intérêts d'apprentissage des chercheurs et des acteurs de l'action publique. Les intérêts des chercheurs sont privilégiés par cette institution issue du champ scientifique. Les intérêts de connaissance du porteur de projets peuvent être satisfaits uniquement s'ils se calquent sur ceux de l'évaluateur. L'inverse n'a pas été observé. Le détournement de l'institution vers des intérêts propres au porteur de projets engendre une déstabilisation du protocole menaçant la production de résultats. Cependant, il apparaît à l'issue des expérimentations que les acteurs de l'action publique reprennent le contrôle sur les résultats et les interprètent dans un sens leur étant favorable.

Enfin, quel est l'impact de cette institution d'apprentissage scientifique sur l'apprentissage des acteurs de l'action publique ? Nous comprenons ici aussi bien l'apprentissage pendant la conduite de l'expérimentation que celui issu des résultats produits par l'expérimentation.

Nous avons émis l'hypothèse que le processus de production de connaissance de la méthode aléatoire constituait un obstacle à l'apprentissage des acteurs de l'action publique pendant la conduite de l'expérimentation. Nos études de cas confirment que ces deux logiques sont en opposition mais témoignent davantage d'un contournement de cette obligation par les porteurs de projets ayant contribué à menacer le processus de production des résultats scientifiques. Il apparaît que même dans un cadre extrêmement structuré, les acteurs de la mise en œuvre de l'action publique ont continué leurs pratiques d'ajustement, et d'interprétation de l'intervention, s'inscrivant dans leur intérêt selon leur position dans leur propre *champ d'action stratégique*. Ce *travail institutionnel* de déstabilisation de l'institution expérimentale varie en fonction de l'atteinte du contrôle expérimental sur les pratiques et la culture professionnelle des acteurs de terrain. Ce travail peut être volontaire et argumenté par des individus questionnant la pertinence de ce contrôle sur leur domaine d'intervention ou être aussi uniquement le fruit de pratiques routinières de l'action publique. Tout dispositif faisant appel à une interaction entre l'opérateur et le bénéficiaire comportait une part de

subjectivité et a donc évolué dans le temps et varié de sites en sites. Les interventions ont été ajustées au fur et à mesure de leur mise en œuvre au gré des apprentissages des acteurs de terrain. Ce phénomène amène à questionner la pertinence de l'application de cette méthode d'évaluation à des interventions innovantes fondées sur l'action d'opérateurs de terrain au vu des risques d'écarts trop importants entre le traitement prévu à l'origine et son incarnation sur le terrain. Pour une production rigoureuse de résultats, cette méthode nécessite une stabilité du traitement peu présente dans l'intervention publique.

Enfin, nous souhaitons aussi identifier le modèle d'apprentissage proposé par la méthode expérimentale par assignation aléatoire et le situer dans les différentes *espèces* d'apprentissages du *genre* « apprentissage épistémologique » proposées par C. Radaelli et C. Dunlop. Pour rappel, ces catégories étaient organisées suivant une matrice à deux variables, le contrôle de l'apprenant sur les objectifs et la finalité de l'apprentissage d'une part et sur le contenu et les moyens de l'apprentissage d'autre part. La particularité de la méthode aléatoire est que les formateurs recherchent aussi un apprentissage dans cette activité. Dans une relation évaluative typique, les évaluateurs interviennent comme des *facilitateurs*, c'est-à-dire que les acteurs de l'action publique définissent les objectifs d'apprentissage et que les évaluateurs interviennent uniquement avec des méthodologies de recherche ou d'enquête pour accompagner cet apprentissage. Dans les relations cliniques que nous avons étudiées, le contrôle nécessaire de l'action publique par cette méthode, pose les chercheurs/évaluateurs en *enseignants*. En d'autres termes, ils souhaitent contrôler autant la fin que les moyens d'apprentissage puisque cette méthode ne peut s'intéresser qu'à un impact précis préalablement défini et que la rigidité de son protocole nécessite une maîtrise du processus de production de connaissances. Aussi, à l'issue de cette relation clinique, il semble que finalement l'évaluateur est relégué à une fonction de *contributeur*, c'est-à-dire que son travail est considéré comme un élément d'information parmi d'autres, et que les acteurs de l'action publique prennent leurs décisions selon leur propres apprentissages qui ne sont pas seulement issus des résultats de l'évaluation, de leurs interprétations et leurs objectifs. Ainsi, il apparaît que le travail institutionnel au sein d'un même instrument d'apprentissage peut amener ce dernier à dériver entre plusieurs *espèces* d'apprentissage à différents moments de son utilisation. La méthode expérimentale par assignation aléatoire, bien qu'elle ait l'ambition de cadrer rigoureusement la production de connaissance, ne structure pas les modalités d'apprentissage des acteurs de l'action publique puisque qu'elle n'a pas vocation à organiser un apprentissage pendant sa conduite (où l'apprentissage scientifique est privilégié) ni lors de la production des résultats (puisque beaucoup d'évaluations n'apportent pas de données suffisantes ou dans les délais suffisant pour éclairer la décision et qu'une perspective cumulative de connaissance amène régulièrement ces praticiens à rester réservés dans leur recommandations). Il semble ainsi qu'à défaut d'être un instrument favorisant l'apprentissage politique par l'administration de la preuve, la méthode expérimentale par assignation aléatoire se présente davantage comme un instrument visant à organiser la mise en œuvre de l'action publique autour des standards de la production de résultats scientifiques expérimentaux.

Conclusion

A l'issue de ce travail de recherche, plusieurs enseignements se dégagent sur l'utilisation de la méthode expérimentale par assignation aléatoire et plus généralement sur les instruments d'action publique ayant pour visée d'organiser un apprentissage politique par des méthodes scientifiques. S'il est encore prématuré de se prononcer sur le processus d'institutionnalisation de cette méthode en France, il est néanmoins possible d'identifier des tendances stables sur les modalités d'interactions structurées par cette méthode entre les acteurs scientifiques et ceux de l'action publique. Nous présentons d'abord, les apports de notre appareillage théorique afin de comprendre les modalités de la constitution d'un champ d'action stratégique et de sa dissolution autour d'un projet d'institutionnalisation d'un nouvel instrument. Nous revenons ensuite, sur les variables explicatives, mises à jour par nos hypothèses de recherche, articulant les pratiques, les intérêts, au sein de cette interaction et qui structurent les modalités d'incarnation de cette méthode. Nous dégageons notamment des enseignements sur les modalités d'apprentissages politiques organisés par la méthode expérimentale par assignation aléatoire lorsqu'elle est appliquée à des innovations sociales.

Un instrument d'apprentissage renégocié par des interactions multi-niveaux

Cette étude empirique de la diffusion et de la pratique de l'expérimentation aléatoire montre que cet instrument, qui semble se distinguer de prime abord par sa rigidité, sa neutralité et sa scientificité, est constamment l'objet de renégociations au gré autant des conjonctures, que des intérêts, des pratiques et de l'interaction des acteurs qui s'en saisissent. Cette renégociation s'opère à un double niveau et en deux temps. Elle intervient tout d'abord, lors de la définition de cet instrument comme un outil légitime de l'action publique et de sa tentative d'institutionnalisation, ce que nous avons étudié dans le chapitre 2. Elle se rejoue ensuite localement lors de la mise en œuvre effective de celui-ci par les acteurs, ce que nous avons étudié dans le chapitre 3.

Les stratégies d'insitutionnalisation, l'échec de la stabilisation d'un champ d'action stratégique

Notre étude de la tentative d'institutionnalisation de la méthode expérimentale au sein de l'administration centrale en France, nous a permis d'observer les difficultés liées à ce processus de renégociation. Les concepts de *champs d'action stratégique* et de *travail institutionnel* ont été particulièrement féconds pour rendre compte de ce phénomène. Il apparaît d'abord, que les variables explicatives de l'émergence d'un nouvel instrument tiennent à l'intérêt du positionnement qu'il est susceptible d'offrir à ses promoteurs. Les promoteurs ou *entrepreneurs* se structurent alors en champ d'action stratégique émergent dont l'objet est principalement l'institutionnalisation de cet instrument, dont ils seront les *titulaires*. De plus, cet instrument constitue une ressource qu'ils peuvent importer dans leur champ d'action stratégique traditionnel d'appartenance afin d'améliorer ou de consolider leur position voire leur notoriété. Ces entrepreneurs choisissent donc de promouvoir un instrument au plus près de leur *pratique* afin de bénéficier d'un avantage sur leurs potentiels *challengeurs*.

Cependant, cette entreprise nécessite aussi un travail institutionnel de mobilisation de tout un ensemble d'autres acteurs afin de la soutenir. Elle requiert donc d'ajuster l'instrument, tel un *objet frontière* selon les groupes de personnes à mobiliser. Cet instrument est donc sujet à renégociation ou encore à redéfinition, afin de présenter la facette la plus avantageuse pour le groupe à mobiliser et la conjoncture nationale. Cette

entreprise de renégociation de la raison d'être de l'instrument ou de ses composants, peut amener à l'éloigner de son modèle original qui avait suscité l'intérêt de ses promoteurs. En effet, si ces nouveaux groupes sont mieux positionnés dans les champs d'action stratégiques ayant une relation hiérarchique avec le nouveau champ émergent, ils peuvent détourner cette entreprise en leur faveur.

Ainsi, alors que le champ émergent de l'expérimentation sociale avait été défini par ses entrepreneurs comme l'alliance d'un dispositif innovant et de son évaluation par la méthode aléatoire, il a été détourné par les membres des champs d'action stratégiques étatiques pouvant intervenir dans les unités de gouvernance internes du champ. Le Fonds d'Expérimentation pour la Jeunesse qui devait constituer l'organisation d'institutionnalisation de cette méthode d'évaluation, a accordé de moins en moins de place à celle-ci dans ses financements lorsqu'il a été récupéré par d'autres acteurs. Cette étude montre combien, dans les entreprises d'institutionnalisation, l'importance de *suiveurs* au-delà des *entrepreneurs*. Bien que le travail des entrepreneurs, de nature cognitive, organisationnelle, de captation de ressources et de modification de la norme, soit fondamental pour amorcer le processus d'émergence du champ d'action stratégique, ce dernier ne peut se maintenir sans un travail institutionnel, conduit par un cercle d'acteurs plus important. Pour mobiliser ce cercle de *suiveurs* il faut, d'une part, qu'ils trouvent un intérêt à soutenir cette institution mais aussi, d'autre part, que celle-ci ne développe pas des pratiques trop éloignées de leurs pratiques de champ habituelles.

Ainsi, les entrepreneurs doivent atteindre un équilibre fragile, puisqu'ils promeuvent un instrument leur offrant une position privilégiée mais qui doit correspondre aussi aux intérêts et pratiques d'un assez grand nombre d'organisations ou d'individus pour qu'ils le soutiennent. Si ce champ d'action stratégique soutenant cet instrument parvient à susciter l'intérêt de nouveaux individus mais s'écarte trop fortement de leurs *pratiques*, ils risquent de devenir des *challengeurs* qui, selon leur positionnement dans les champs d'action stratégiques interreliés à ce dernier, peuvent s'engager dans un travail de déstabilisation ou de détournement à leur avantage. Par exemple, pour notre objet d'étude, la dimension de financement d'innovations sociales de terrain a été une stratégie d'intéressement efficace des acteurs de l'action publique et le financement d'évaluation « rigoureuses » a aussi mobilisé l'intérêt d'évaluateurs. Cependant la particularité et la distance de la pratique de l'expérimentation aléatoire vis-à-vis de leur pratiques habituelles est devenu un obstacle à son adoption. Le constat de cet échec conduit les entrepreneurs à cesser l'entreprise d'intéressement afin de maîtriser l'intégrité de l'instrument selon leurs intérêts.

Cette tendance se poursuit avec le changement de majorité présidentiel. Cette alternance politique nous offre une fenêtre d'observation sur l'accès des anciens promoteurs de cette méthode dans le contexte d'une nouvelle équipe politique. Le Fonds d'Expérimentation pour la Jeunesse existe toujours. Il publie environ un appel à projets par an, ainsi que des synthèses récapitulatives des résultats d'évaluations⁹⁹³. Cependant, l'écart entre le conseil de gestion et le conseil scientifique continue de se creuser, comme semble en témoigner la démission du président de ce dernier. Si les anciens promoteurs de cette méthode ne bénéficient plus d'une influence particulière sur cette organisation, ils restent des individus détenteurs d'un accès privilégié à l'administration centrale. Ainsi, un travail disparate de promotion de la méthode aléatoire se poursuit. L'objectif n'est plus d'élargir le cercle d'acteurs scientifiques la pratiquant, mais plutôt

⁹⁹³ FONDS D'EXPERIMENTATION POUR LA JEUNESSE, *Premiers enseignements des expérimentations, synthèses thématiques 2013*, Ministère des Sports, de la Jeunesse, de l'Education populaire et de la Vie associative, 2013.

de permettre aux chercheurs, déjà experts, de conduire cette méthode sur de nouveaux terrains ou encore de consolider ou d'affiner des résultats précédents.

De nouvelles instances dédiées en partie à l'évaluation dans l'administration centrale et rattachées au premier ministre, ont été créées comme le Commissariat Général à la Stratégie et la Prospective (CGSP) - qui remplace le Centre d'analyse stratégique et le conseil de l'emploi des revenus et de la cohésion sociale – et le Secrétariat Général pour la Modernisation de l'Action Publique qui selon l'ancien premier ministre (J-M Ayrault) permettra que « toutes les politiques publiques, sur l'ensemble du quinquennat, feront l'objet d'une évaluation »⁹⁹⁴. La logique des évaluations conduites dans le cadre de la MAP (Modernisation de l'Action Publique) est présentée comme « constructiviste », « intersubjective » où l'utilité de l'évaluation prime sur sa scientificité si un arbitrage est nécessaire⁹⁹⁵. L'expérimentation au sens d'innovation sociale est toujours envisagée mais elle n'a pas recours systématiquement à l'évaluation par la méthode aléatoire. Celle-ci demeure toujours présente dans le paysage des méthodes d'évaluation mais reste limitée à un nombre restreint de projets et conduite par les anciens promoteurs de cette méthode. Toutefois, ces derniers bénéficient toujours, d'un accès privilégié à l'administration centrale comme en témoigne une note du Conseil d'Analyse Economique de 2013, présentant l'expérimentation aléatoire comme la méthode la plus rigoureuse, et désignant les chercheurs comme les experts techniques les plus compétents en évaluation⁹⁹⁶. Par ailleurs, certains de ces chercheurs comme Y. L'horty, un des promoteurs clés de la méthode aléatoire en France, siège toujours dans les instances d'évaluation des Ministères comme le comité national d'évaluation des territoires d'excellence du Ministère des droits des Femmes. En parallèle, les anciens membres du Cabinet de M. Hirsch continuent à promouvoir le recours à cette méthode quand ils en ont l'opportunité. Ils sont présents dans le Cabinet du Ministère des droits des femmes – E. Grass en est le directeur –, ou au Ministère des affaires sociales et de la Santé – J-B Dujol en était le sous-directeur de l'accès aux soins, des prestations familiales et des accidents du travail à la Direction de la Sécurité Sociale -⁹⁹⁷. Maintenant qu'il est devenu, depuis mai 2014, Directeur de la jeunesse, de l'éducation populaire et de la vie associative, délégué interministériel à la jeunesse au ministère des Droits des femmes, de la Ville, de la Jeunesse et des Sports, on peut se demander si une nouvelle impulsion pourra être donnée à cette méthode dans le secteur de la jeunesse. M. Hirsch quant à lui est maintenant retourné dans le secteur de la santé, il occupe le poste de directeur général de l'Assistance Publique-Hôpitaux de Paris.

Enfin, les chercheurs ayant recours à cette méthode poursuivent la consolidation s'instances leur permettant de se regrouper pour constituer des équipes d'évaluation au service de l'action publique. On note notamment la constitution de l'IPP (l'Institut des Politiques Publiques) qui réunit les chercheurs économétriciens intéressés par l'évaluation et entérine le partenariat entre les chercheurs du CREST et de l'Ecole d'Economie de Paris. Cet institut organise la valorisation des travaux sur l'action publique ainsi que leur traduction pour en favoriser l'utilisation, et leur permet plus

⁹⁹⁴ Déclaration de Jean-Marc Ayrault au Comité interministériel pour la modernisation de l'action publique, 18 décembre 2012

⁹⁹⁵ Olivier PICAUVET, *Expérimenter ou évaluer les politiques publiques*, séminaire praticiens/chercheurs, ICPDE/SGMAP, 16 novembre 2013.

⁹⁹⁶ LES MEMBRES DU CONSEIL D'ANALYSE ECONOMIQUE, *Evaluation des politiques publiques*, Les notes du conseil d'analyse économique, n°1, Conseil d'analyse économique, 2013. Cette note fut l'objet d'une lettre officielle du président de la Société Française d'Evaluation précisant que d'autres méthodes non expérimentales ou quasi-expérimentales pouvaient aussi permettre d'estimer un impact et que les cabinets d'évaluation privés pouvaient aussi se positionner comme des experts techniques de l'évaluation indépendants.

⁹⁹⁷ Ce dernier explique dans un entretien : « On est en train de négocier la Convention d'Objectifs et de Gestion, le contrat qui lie l'état à la Cnaf et il y a un volet accès au droit, j'ai mis une phrase codée après en avoir discuté avec Marc Gurgand qui autorise, qui demande à la Cnaf si on peut faire des expérimentations sur le recours au RSA »

aisément de répondre ensemble à des appels d'offres. La méthode expérimentale par assignation aléatoire est une des méthodes proposées par cette structure. Par ailleurs, le recours à l'innovation sociale est aussi promu par les institutions européennes qui, par un guide, à destination des acteurs publics, reconnaît aussi la méthode expérimentale par assignation aléatoire comme étalon or des méthodes d'évaluation d'impact⁹⁹⁸. Les chercheurs français souhaitant utiliser cette méthode se tournent donc vers l'évaluation de ces projets qui leurs garantissent de plus grands échantillons. Dans la programmation 2014-2020 le programme Progress du futur Programme européen pour le changement social et l'innovation sociale a eu 17% de son budget consacrés à la promotion de l'expérimentation sociale « *en tant que méthode d'essai et d'évaluation de solutions innovantes, en vue de les appliquer plus largement* »⁹⁹⁹. Ainsi, bien que nous ne puissions conclure à une institutionnalisation de cette méthode d'évaluation au sein de l'administration centrale française, il semble cependant, que sa promotion ait toujours cours pour l'ériger comme un idéal de rigueur d'estimation d'impact.

Une institution d'apprentissage en tension : l'équilibre fragile d'une relation clinique à trois parties

Comme nous l'avons étudié dans notre 3^{ème} chapitre, bien que cette méthode n'ait pas été institutionnalisée à la hauteur des espérances de ses promoteurs, leurs initiatives ont permis de faire émerger plusieurs incarnations pratiques de cet instrument. L'application de cette méthode a été le théâtre de nouvelles négociations entre les acteurs y participant. Plusieurs *sites institutionnels* ont émergé des interactions organisées par cette méthode dont les équilibres ont variés selon les pratiques et les intérêts des parties prenantes. Nous avons démontré que ces intérêts et pratiques étaient eux-mêmes dépendants de la position des acteurs dans leurs champs d'action stratégiques respectifs. Malgré la rigidité apparente de cette méthode, il apparaît que ces sites institutionnels sont soumis au travail institutionnel de maintien et de déstabilisation des acteurs qui la composent. Cette institution est particulièrement sujette au travail institutionnel des acteurs de l'action publique, et notamment ceux concernés par la mise en œuvre, puisqu'elle leur impose des pratiques radicalement différentes des leurs, comme notamment l'allocation du traitement aux bénéficiaires par assignation aléatoire ainsi que la stabilité dans le temps et dans l'espace de l'administration de ce dernier.

L'ambition, défendue par les promoteurs de cette méthode, d'organiser une relation clinique où chacun réalise son objectif d'apprentissage grâce à un instrument commun semble difficile à atteindre. L'expérimentation aléatoire devient peu à peu pour les chercheurs/évaluateurs, qui étaient aussi les promoteurs de cette méthode, une pratique routinisée dont ils apprennent à maîtriser les rouages afin de garantir la production de connaissances pertinentes pour leur recherche et valorisables dans le champ d'action stratégique international de l'économie appliquée. La production de ces résultats a été malgré tout, dans un premier temps, fortement perturbée par le processus d'apprentissage des acteurs de l'action publique participant à un travail institutionnel de

⁹⁹⁸ On lit dans ce guide: "*Random assignment to treatment is regarded as the "gold standard" for constructing a valid comparison group. But as we will emphasize in the next part of this guide, randomized evaluations of social programs take time and can be complex to implement. For these reasons, other techniques are also commonly used. They are referred to as non-experimental or quasi-experimental methods. They are usually less complex to implement than randomized evaluations, but the results they deliver are also less reliable*". J-PAL EUROPE, *Social Experimentation, A methodological guide for policy makers*, Directorate General for Employment, Social Affairs and Inclusion, European Commission, 2011.

⁹⁹⁹ Pour un retour détaillé sur le développement de l'innovation sociale et du recours à la méthode expérimentale par assignation aléatoire par les institutions de l'Union européenne voir l'article : Marjorie JOUEN, *Les expérimentations sociales en Europe, Vers une palette plus complète et efficace de l'action communautaire en faveur de l'innovation sociale*, Notre Europe, 2008.

déstabilisation. Des éléments propres à l'apprentissage des acteurs de l'action publique lors de projets innovants, comme la redéfinition des publics cibles, des moyens de les mobiliser, l'appropriation du dispositif par les opérateurs de terrain, ainsi que son ajustement, sont autant d'éléments qui contribuent au travail institutionnel de déstabilisation de cette institution scientifique. On observe donc, une tension entre ces deux pratiques d'apprentissage. L'analyse de nos études de cas montre que chacun de ces apprentissages – des acteurs scientifiques ou de ceux de l'action publique – ne s'organisent qu'au détriment l'un de l'autre.

Ces tensions sont révélatrices de l'ambition de réorganisation de la nature de l'interaction entre acteurs de l'action publique et acteurs scientifiques portée par cette méthode. En effet, plus la conduite de cette méthode correspond aux standards de la production de connaissances scientifiques, plus elle semble s'éloigner des conditions d'apprentissage des acteurs de l'action publique portant le dispositif évalué. Elle crée une situation de test artificiel qui ne permet pas de susciter des apprentissages sur une action publique conduite dans un environnement ouvert. Cette tension est d'ailleurs inhérente au terme « expérimenter » qui désigne le fait de « pratiquer des opérations destinées à étudier quelque chose » par un protocole de démonstration scientifique, mais aussi de « connaître par expérience personnelle », de manière beaucoup plus subjective et souvent par tâtonnement. Si le premier sens correspond à l'acception scientifique du terme, le second semble se référer à la conception des acteurs de l'action publique de l'apprentissage expérimental. Le couplage de ces deux logiques d'apprentissage apparaît difficile, puisque justement l'expérimentation scientifique tend vers le contrôle de l'appréciation et de l'ajustement subjectif des acteurs de terrain, au regard de leur expérience. Lorsque les acteurs de l'action publique s'engagent dans ce type d'apprentissage, c'est alors les conditions de production scientifiques qui sont fragilisées.

L'impact de cette méthode sur l'action publique semble ainsi surtout porter sur la conception des projets expérimentaux et les modalités de leur mise en œuvre, impactant de fait la conduite de l'action publique sujette à évaluation. Le fait que cette méthode soit conduite par des économistes statisticiens poursuivant un objectif de production scientifique, les amène à se situer comme co-expérimentateurs, utilisant le terrain de l'expérimentation sociale pour tester des théories de micro-économie comportementaliste. Ils favorisent donc l'élaboration de projets proposant une intervention simple, standardisée, ponctuelle, ne fonctionnant pas de manière systémique avec d'autres dispositifs, et basée souvent sur la réaction de publics à un type d'incitation. D'ailleurs, comme en témoignent les rapports d'évaluation et les publications scientifiques, l'expérimentation aléatoire n'est pas une institution d'apprentissage visant l'apprentissage des acteurs de l'action publique portant ou conduisant le projet évalué. La perspective cumulative de création de connaissance scientifique qu'elle structure l'amène à viser un apprentissage conceptuel sur des mécanismes d'action publique. Par exemple, elle peut générer de la connaissance scientifique transversale à plusieurs dispositifs sur le mécanisme de recours aux contrats dans l'accompagnement des publics en difficultés. Ces enseignements s'adressent donc à une entité politique, seulement mobilisable à plus long termes, lorsque les résultats seront stabilisés. Elle n'organise donc pas une relation clinique réciproque lors de l'expérimentation mais un apprentissage scientifique, cumulatif qui à terme pourrait produire des résultats susceptibles d'être saisis par les décideurs publics. Néanmoins, à ce jour, même dans le cadre de l'imposition de cette utilisation aux Etats-Unis, celle-ci semble toujours principalement guidée par des variables liées à l'intérêt politique du moment et allant souvent à l'encontre des règles de généralisation de résultats

scientifiques¹⁰⁰⁰.

Si cette méthode est institutionnalisée comme instrument privilégié d'une démarche *d'evidence-based policy*, on observe alors une dynamique récursive d'influence réciproque entre la production scientifique et l'action publique. Cette dernière, dans cette conjoncture, interfère à la fois sur les modalités de production scientifique et les débats épistémologiques, puisqu'elle désigne une méthode officielle de démonstration et privilégie son financement, ce qui a pour effet de délégitimer la scientificité des autres méthodes de démonstration. En retour, cette méthode, chargée d'une théorie de démonstration de la causalité et de présupposés sur le comportement des individus, tend à modéliser les interventions auxquelles elle s'applique selon les impératifs de sa production de connaissance. Si l'exigence d'être évalué par cette méthode est poussée à l'extrême, comme nous l'avons vu pour le ministère de l'éducation aux Etats-Unis, cela peut conduire à accorder des financements uniquement aux projets correspondants à ses critères d'évaluabilité et à écarter les autres.

Ces résultats invitent cependant à se demander, si tout dispositif évaluatif, notamment lorsqu'il est élaboré en amont d'une innovation sociale n'a pas toujours un impact significatif sur celle-ci ? D'autant plus, lorsqu'elle est conduite par des chercheurs qui tendent à structurer l'intervention pour qu'elle réponde à leurs objectifs de production de connaissances parfois au détriment de l'apprentissage des acteurs de terrain ? Puisque nous avons nous même occupé la position d'évaluatrice d'innovations sociales, sans avoir recours à la méthode aléatoire, ce questionnement est aussi réflexif sur notre propre pratique.

Selon nos observations et notre pratique évaluative, il est clair que tout dispositif d'évaluation échafaudé en amont d'un projet, en structure la mise en œuvre et les objectifs. L'intérêt de l'étude de la méthode expérimentale par assignation aléatoire est qu'elle rend manifeste cette structuration par l'exigence de la conformité de l'action à sa méthode de démonstration de la causalité. La rigueur du protocole qu'elle nécessite, révèle en creux des dynamiques à l'œuvre dans les projets innovants de l'action publique, entrant en opposition avec cette pratique scientifique. Néanmoins, il semble qu'une approche de recherche évaluative par abduction ouvre un espace d'interaction des sphères scientifiques et de l'action publique laissant une place à l'apprentissage de ces deux groupes d'acteurs. Cette démarche implique un mouvement itératif entre l'incarnation du projet et l'appareillage scientifique pour l'appréhender ainsi que leur évolution mutuelle au cours de la mise en œuvre du projet. Appelée « *developmental evaluation* », cette approche évaluative consacrée à l'innovation sociale commence à faire ses preuves aux Etats-Unis¹⁰⁰¹. Cependant, elle pose la question fondamentale du destinataire de l'apprentissage évaluatif. Cette démarche centre le travail évaluatif sur l'apprentissage des acteurs de terrain et des porteurs de projets, et ne peut lors de cette phase d'apprentissage viser à rendre compte quantitativement de l'impact de ce projet à un autre interlocuteur. Mais l'innovation sociale, n'a-t-elle pas vocation intrinsèquement à constituer un moment de répit dans l'exigence de résultats jusqu'à ce qu'elle se stabilise ? L'intelligence collective dont-elle est le fruit ne peut-elle pas être le garant, pour une période donnée, du souci d'efficacité de cette intervention ? Le chercheur/évaluateur peut alors prendre un rôle de facilitateur de l'apprentissage des

¹⁰⁰⁰ L'article de C. Weiss sur *l'imposed use* aux Etats-Unis revient précisément sur l'absence de critères de scientificité dans ce type de prise de décision Carol H. WEISS, « The Fairy Godmother and Her Warts: Making the Dream of Evidence-Based Policy Come True », (2008) 29-29 *American Journal of Evaluation*.

¹⁰⁰¹ Nous renvoyons le lecteur intéressé au passionnant ouvrage de M; Q. Patton présentant les fondements théoriques et des cas d'application de cette pratique évaluative Michael Quinn PATTON, *Developmental evaluation : applying complexity concepts to enhance innovation and use*, New York, Guilford Press, 2010.

acteurs de l'action publique, à l'aide de ses outils d'analyse et de collecte de données ainsi que de sa posture critique. Il doit cependant abandonner toute ambition d'organiser une production de connaissance suivant des protocoles, des hypothèses, et des objectifs fixés *a priori* et destinés à rester stables dans le temps. Cette posture de recherche nécessite alors de ne pas chercher à contrôler l'action publique mais de se donner les moyens de l'appréhender dans toute sa complexité.

Bibliographie

Ouvrages

- ALKIN M., CHRISTIE C., "An evaluation theory tree revisited", in ALKIN M. C. (dir.), *Evaluation roots: a wider perspective of theorists' views and influences*, 2nd ed, Los Angeles, SAGE Publications, 2004.
- AMANN R., "Foreword", in NUTLEY S., SMITH P., DAVIES H., *What works?: evidence-based policy and practice in public services*, Bristol, The Policy Press, 2000.
- ARENDT H., « Vérité et politique », in *La Crise de la culture: huit exercices de pensée politique*, Paris, Gallimard, 1989.
- ARGYRIS C., SCHÖN D. A., *A theory of action perspective*, Reading, Mass, Addison-Wesley Pub. Co, 1978.
- ATTEN M., PESTRE D., *Heinrich Hertz: l'administration de la preuve*, Paris, Presses Universitaires de France, 2002.
- BACHELARD G., *La formation de l'esprit scientifique: contribution à une psychanalyse de la connaissance*, Paris, J. Vrin, 1934.
- BACHIR M., BUSSAT V., « L'entretien en actes », in BACHIR M. (dir.), *Les méthodes au concret: démarches, formes de l'expérience et terrains d'investigation en science politique*, Presses Universitaires de France, 2000, p. 31-58.
- BECKER H., *Écrire les sciences sociales*, Paris, Economica, 2004.
- _____, *Les ficelles du métier*, coll. Guide repère, La Découverte, 2002.
- BERGER P., LUCKMANN T., *La construction sociale de la réalité*, Paris, Armand Colin, 2006.
- BERTHELOT J.-M., *L'emprise du vrai: connaissance scientifique et modernité*, Paris, Presses Universitaires de France, 2008.
- BHASKAR R., *A realist theory of science*, coll. Radical thinkers, n°29, London ; New York, Verso, 2008.
- BOUSSAGUET L., JACQUOT S., RAVINET P., *Dictionnaire des politiques publiques*, Paris, Presses de Sciences Po, 2010.
- BRÉGEON P., *À quoi servent les professionnels de l'insertion? L'exemple des intervenants sociaux en centres d'hébergement et de réinsertion sociale, en missions locales et dans l'insertion par l'activité économique*, coll. Le travail du social, Paris, L'Harmattan, 2008.
- CAMPBELL D. T., STANLEY J. C., *Experimental and Quasi-Experimental Design for Research*, Boston, Houghton Mifflin, 1963.
- CONSEIL SCIENTIFIQUE DE L'ÉVALUATION, *Petit guide de l'évaluation des politiques publiques*, Paris, La Documentation française, 1996.
- CRONBACH L. J. (dir.), *Toward reform of program evaluation*, 1st ed, coll. A Joint publication in the Jossey-Bass series in social and behavioral science & in higher education, San Francisco, Jossey-Bass Publishers, 1980.
- DENZIN N., *Sociological Methods*, New York, McGraw-Hill, 1978.
- DESROSIERES A., La statistique, outil de gouvernement et outil de preuve, dans *L'argument statistique*, Paris, Presses de l'école des mines, 2008, en ligne : http://www.meshs.fr/documents/pdf/seminaires/2009/desrosieres_1_09.pdf (consulté le 15 juillet 2012).
- _____, Statistique et espace public: le rêve de l'extériorité, dans *Le savant et le politique aujourd'hui*, Bibliothèque Albin Michel, 1996, p. 107-118.

-
- L'argument statistique II, *Gouverner par les nombres*, coll. Sciences sociales, Paris, Presses de l'école des mines, 2008.
- DIMAGGIO P., « Interest and agency in institutional theory », in ZUCKER L. G. (dir.), *Institutional patterns and organizations: Culture and environment*, Cambridge, Mass, Ballinger, 1988, p. 3-22.
- DOBRY M., *Sociologie des crises politiques la dynamique des mobilisations multisectorielles*, 3e édition, Paris, Presses de Science Po, 2009.
- DONALDSON S., LIPSEY M., « Roles for theory in Contemporary Evaluation Practice: Developing Practical Knowledge », in *Handbook of evaluation*, 2006.
- DONALDSON, S. I., CHRISTIE C., MARK M. (dir.), *What counts as credible evidence in applied research and evaluation practice?*, Los Angeles, SAGE, 2009.
- DONEGANI J.-M., DUCHESNE S., HAEGEL F., « Sur l'interprétation des entretiens de recherche », in *Aux frontières des attitudes entre le politique et le religieux: textes en hommage à Guy Michelat*, coll. Collection Logiques politiques, Paris, L'Harmattan, 2002, p. 273-295.
- DROITCOUR J., KOVAR M., « Multiple Threats to the Validity of Randomized Studies », in SMITH N. L., BRANDON P. R. (dir.), *Fundamental issues in Evaluation*, The Guildford Press, New York, 2008, p. 65-88.
- DRYZEK J. S., « Policy Analysis and Planning: From Science to Argument », in FISCHER F., FORESTER J., *The Argumentative Turn in Policy Analysis and Planning*, Taylor & Francis, 1993, p. 213-232.
- DUBOIS V., « Introduction, The sociology of administrative work, a study of "street-level bureaucracy" à la française », in EYMERI-DOUZAN J.-M., BOUCKAERT G., *La France et ses administrations: un état des savoirs*, Bruxelles, Bruylant, 2013, p. 169-177.
- DUBOIS V., *La vie au guichet*, Paris, Économica, 1999.
- DUFOUR S., FORTIN D., HAMEL J., *L'enquête de terrain en sciences sociales: l'approche monographique et les méthodes qualitatives*, Montréal, Éditions Saint-Martin, 1991.
- DUNN W. N., *Public policy analysis: an introduction*, Boston, Pearson, 1994.
- DURAN P., « Genèse de l'analyse des politiques publiques », in RAVINET P., MULLER P., *Dictionnaire des politiques publiques*, Presses de Science Po, Paris, 2006.
- ELLUL J., *L'illusion politique*, Paris, Table ronde, 1965.
- ETHEREDGE L. S., « Government Learning: An Overview », in *The Handbook of Political Behavior* vol. 2, Plenum Press, New York, 1981, p. 73-161.
- FERRACCI M., *Evaluer la formation professionnelle*, Les Presses de Sciences Po, coll. Sécuriser l'emploi, 2013.
- FISHER R. A., *The Design of Experiments*, New York, Hafner Publishing Company, 1935.
- FLIGSTEIN N. MCADAM D., *A theory of fields*, New York, Oxford University Press, 2012.
- FUNNELL S., ROGERS P., « Causal Inference », dans *Purposeful program theory, effective use of theories of change and logic models*, Jossey-bass, 2011, p. 469-499.
- GAXIE D., LABORIER P., « Des obstacles à l'évaluation des actions publiques et quelques pistes pour tenter de les surmonter », in *Etre gouverné: études en l'honneur de Jean Leca*, Paris, Presses de Sciences po, 2003, p. 201-224.
- GEORGE A., BENNETT A., *Case studies and theory development in the social sciences*, Cambridge, Mass, MIT Press, 2005.
- GERSTEN R., HITCHCOCK J., « What is Credible Evidence in Education? The Role of What Works Clearinghouse in Informing the Process », in CHRISTIE C. A., MARK M., DONALDSON S. (dir.), *What counts as credible evidence in applied research and evaluation practice?*, Los Angeles, SAGE, 2009, p. 78-96.
- GIDDENS A., *La constitution de la société: éléments de la théorie de la structuration*, Paris, Presses universitaires de France, 2012.

-
- Central Problems in Social Theory: Action, Structure and Contradictions in Social Analysis, University of California Press, 1979.
- GOFFMAN E., *Asiles; études sur la condition sociale des malades mentaux et autres reclus*, Paris, Editions de minuit, 1968.
- GOODING D., SCHAFFER S., *The Uses of experiment: studies in the natural sciences*, Cambridge [England]; New York, Cambridge University Press, 1989.
- GREENBERG D. H., SHRODER M., *The digest of social experiments*, 3rd ed, Washington, D.C, Urban Institute Press, 2004.
- GUBA E., LINCOLN Y., « The countenances of 4 generation evaluation : Description, judgement and negociation », in PALUMBO D., *The Politics of program evaluation*, coll. Sage yearbooks in politics and public policy, v. 15, Newbury Park, Calif, Sage Publications, 1987, p. 202-234.
- HAMEL J., DUFOUR S., FORTIN D., *Case study methods.*, London, SAGE, 1993.
- HECKMAN J. J., « Randomization and Social Policy Evaluation », in MANSKI C., GARFINKEL I., *Evaluating Welfare and Training Programs*, Harvard University Press, Cambridge, MA, 1992.
- HECLO H., *Modern social politics in Britain and Sweden : from relief to income maintenance*, New Haven [u.a.], Yale Univ. Press, 1975.
- HIRSCH M., *Secrets de fabrication*, Paris, Grasset, 2010.
- HOBBS T., *Léviathan: traité de la matière, de la forme et du pouvoir de la république ecclésiastique et civile*, coll. Classiques des sciences sociales., Chicoutimi, J.-M. Tremblay, 2004, en ligne : http://classiques.uqac.ca/classiques/hobbes_thomas/leviathan/leviathan.html (consulté le 12 mai 2014).
- JEPPERSON R. L., « Institutions, institutional effects and institutionalism », in POWELL W., DiMAGGIO P. (dir.), *The New Institutionalism in Organizational Analysis*, Chicago: University of Chicago Press, 1991, p. 143-163.
- KAHNEMAN D., CLARINARD R., *Système 1, système 2: les deux vitesses de la pensée*, [Paris], Flammarion, 2012.
- LAGROYE J., OFFERLE M., *Sociologie de l'institution*, Paris, Belin, 2010.
- LAPASSADE G., « Observation participante », in ENRIQUEZ E., BARUS-MICHEL J., LEVY A., *Vocabulaire de psychosociologie*, Paris, 2010, p. 375-390.
- LATOUR B., « The force and the reason of experiment », dans LE GRAND HOMER, *Experimental inquiries: historical, philosophical and social studies of experimentation in science*, Dordrecht; Boston; London, Netherlands ; Kluwer academic, 1990.
- LAUFER R., *Le prince bureaucrate: Machiavel au pays du marketing*, Paris, Flammarion, 1982.
- LAWRENCE T. B., SUDDABY R., « Institutions and Institutional Work », in CLEGG S., HARDY C., LAWRENCE T.B., *The Sage handbook of organization studies*, 2nd ed, London ; Thousand Oaks, Calif, Sage Publications, 2006, p. 215-254.
- LE GALES P., LASCOUMES P., « L'action publique saisie par ses instruments », in *Gouverner par les instruments*, Presse de Science po, 2004, p. 11-44.
- LEVY F., « L'évaluation des politiques publiques aux Etats-Unis », in NIOCHE J.-P., POINSARD R., *L'Evaluation des politiques publiques*, Paris, Economica, 1984, p. 27-33.
- LIEBERSON S., « Small N's and big conclusions: an examination of the reasoning in comparative studies based on a small number of cases », in RAGIN C. C., BECKER H., *What is a case? : exploring the foundations of social inquiry*, Cambridge [England]; New York, NY, USA, Cambridge University Press, 1992, p. 105-118.
- LINDBLOM C. E., COHEN D. K., *Usable knowledge: social science and social problem solving*, New Haven, Yale University Press, 1979.

- LIPSKY M., *Street-level bureaucracy: dilemmas of the individual in public services*, New York, Russell Sage Foundation, 2010.
- MARKS H., *La médecine des preuves : histoire et anthropologie des essais cliniques : 1900-1990*, Le Plessis-Robinson, Institut Synthélabo pour le progrès de la connaissance, 1999.
- MEASSON L., MATYJASIK N., « L'évaluation entre opportunité et effet », in FAURE A. (dir.), *Les politiques publiques à l'épreuve de l'action locale: critiques de la territorialisation*, coll. Questions contemporaines, L'Harmattan, 2007.
- MILL J. S., *Système de logique déductive et inductive.*, Chicoutimi, J.-M. Tremblay, 2002, en ligne :
http://classiques.uqac.ca/classiques/Mill_john_stuart/systeme_logique/livre_3/systeme_logique_3.html (consulté le 20 avril 2012).
- MONNIER E., *Evaluations de l'action des pouvoirs publics*, Paris, Economica, 1992.
- NAY O., SMITH A., « Les intermédiaires en politique: médiations et jeux d'institutions », in *Le gouvernement du compromis. Courtiers et généralistes dans l'action politique*, Economica, Paris, 2002, p. 47-86.
- NAY O., *La Région, une institution: la représentation, le pouvoir et la règle dans l'espace régional*, coll. Collection Logiques politiques, Paris, Harmattan, 1997.
- NIOCHE J.-P., POINSARD R., *L'Evaluation des politiques publiques*, Paris, Economica, 1984.
- NORTH D. C., *Institutions, Institutional Change and Economic Performance*, Cambridge University Press, 1990.
- NUTLEY S. M., SMITH P. C., « Evidence and the policy process », in DAVIES H., *What works? : evidence-based policy and practice in public services*, Bristol, The Policy Press, 2000.
- NUTLEY S. M., DAVIES H., TILLEY N., « Debates on the role of experimentation », in NUTLEY S. M., SMITH P., DAVIES H., *What works? : evidence-based policy and practice in public services*, Bristol, The Policy Press, 2000.
- NUTLEY, S. M., SMITH P., DAVIES H., « Introducing evidence-based policy and practice in public services », in *What works? : evidence-based policy and practice in public services*, Bristol, The Policy Press, 2000.
- ORR L., *Social experiments: evaluating public programs with experimental methods*, Thousand Oaks, Calif, Sage Publications, 1999.
- PALUMBO D. J. (dir.), *The Politics of program evaluation*, coll. *Sage yearbooks in politics and public policy*, v. 15, Newbury Park, Calif, Sage Publications, 1987.
- PASSERON J.-C., REVEL J., *Penser par cas*, Paris, Éditions de l'École des Hautes Études en Sciences Sociales, 2005.
- PATTON M. Q., *Developmental evaluation : applying complexity concepts to enhance innovation and use*, New York, Guilford Press, 2010.
- _____, *Utilization-focused evaluation: the new century text*, Thousand Oaks; London; New Delhi, Sage Publications, 1997.
- _____, *Qualitative Evaluation and Research Methods*, Thousand oaks, CA, Sage Publications, 2001.
- PAWSON R., *Evidence-based policy: a realist perspective*, London ; Thousand Oaks, Calif, SAGE, 2006.
- PRESSMAN J. L., WILDAVSKY A. B., *Implementation: how great expectations in Washington are dashed in Oakland : or, why it's amazing that federal programs work at all, this being a saga of the Economic Development Administration as told by two sympathetic observers who seek to build morals on a foundation of.*, Berkeley, Calif., University of California, 1984.
- RIPLEY R. B., *Policy analysis in political science*, coll. The Nelson-Hall series in political science, Chicago, Nelson-Hall, 1985.

- ROSSI P. H., FREEMAN H. E., *Evaluation: a systematic approach*, 5th ed, Thousand Oaks, CA, Sage, 1993.
- SABATIER P., « The political context of evaluation research, an advocacy coalition perspective », in KESSLER M-C., LASCOUMES P., THOENIG J-C., *Evaluation des politiques publiques*, coll. Collection Logiques politiques, Paris, Harmattan, 1998, p. 129-146.
- SALSBERG D., *The lady tasting tea: how statistics revolutionized science in the twentieth century*, New York, W.H. Freeman, 2001.
- SCHOLZ R. W., TIETJE O., *Embedded case study methods: integrating quantitative and qualitative knowledge*, Thousand Oaks, Calif, Sage Publications, 2002.
- SCOTT W. R., MEYER J. W., « The organization of societal sectors; proposition and early evidence », in POWELL W., DIMAGGIO P., *The New Institutionalism in Organizational Analysis*, Chicago: University of Chicago Press, 1991, p. 109-142.
- SCOTT W. R., *Institutions and organizations*, Thousand Oaks, California, Sage publication, 1995.
- SCRIVEN M., « Demythologizing Causation and Evidence », in DONALDSON S., CHRISTIE C., MARK M. M (dir.), *What Counts as Credible Evidence in Applied Research and Evaluation Practice?*, Sage, 2009.
- SHADISH W. R., LUELLEN J. K., « Donald Campbell: The Accidental Evaluator », in ALKIN M. C. (dir.), *Evaluation roots: a wider perspective of theorists' views and influences*, 2nd ed, Los Angeles, SAGE Publications, 2004, p. 80-87.
- SHADISH W. R., COOK T. D., LEVITON L. C., « Social Program Evaluation: Its History, Tasks, and Theory », in *Foundations of program evaluation: theories of practice*, Newbury Park, Calif., Sage Publications, 1991, p. 19-35.
- SHADISH W. R., COOK T. D., LEVITON L. C., *Foundations of program evaluation: theories of practice*, Newbury Park, Calif., Sage Publications, 1991.
- SHAPIN S., SCHAFFER S., *Leviathan and the air-pump : Hobbes, Boyle, and the experimental life*, Princeton, NJ, Princeton University Press, 1985.
- SMITH N. L., BRANDON P. R (dir.), *Fundamental issues in evaluation*, New York, Guilford Press, 2008.
- STAME N., « Avantages et inconvénients des différentes méthodes d'évaluation: comment choisir? », in TROSA S. (dir.), *Évaluer les politiques publiques pour améliorer l'action publique une perspective internationale*, Paris, Comité pour l'histoire économique et financière de la France, 2009.
- STEINMO S., THELEN K., « Historical institutionalism in Comparative Analysis », dans STEINMO S., THELEN K. A., LONGSTRETH F., *Structuring politics: historical institutionalism in comparative analysis*, coll. Cambridge studies in comparative politics, Cambridge [England] ; New York, Cambridge University Press, 1992.
- TAYLOR S. J., BOGDAN R., *Introduction to qualitative research methods : a guidebook and resource*, New York, Wiley, 1998.
- TOLBERT P. S., ZUCKER L. G., « The institutionalization of institutional theory », in CLEGG S., HARDY C. (dir.), *Handbook of organization studies*, Sage, London, 1996, p. 175-190.
- TOPALOV C., « Le champ réformateur, 1880-1914: un modèle », dans *Laboratoires du nouveau siècle: la nébuleuse réformatrice et ses réseaux en France, 1880-1914*, coll. Civilisations et sociétés, n°98, Paris, Ecole des hautes études en sciences sociales, 1999.
- WACQUANT L., *Corps & âme : carnets ethnographiques d'un apprenti boxeur*, Marseille, Agone, 2002.
- WEBER M., *Le savant et le politique*, Paris, Plon, 2002.

- WEISS C. H., « Evaluation Research for Decision: Is Anybody There? Does Anybody Care? », in ALKIN M. C., *Debates on evaluation*, Newbury Park, Calif., Sage Publications, 1990.
- WEISS C. H., *Evaluation research: methods for assessing program effectiveness*, coll. Prentice-Hall methods of social science series, Englewood Cliffs, N.J, Prentice-Hall, 1972.
- WIEVIORKA M., « Case studies: history or sociology? », in RAGIN C. C., BECKER H., *What is a case? : exploring the foundations of social inquiry*, Cambridge [England]; New York, NY, USA, Cambridge University Press, 1992.
- WILDAVSKY A. B., *Speaking truth to power : the art and craft of policy analysis*, New Brunswick, USA, Transaction Books, 1987.
- YIN R. K., *Case study research : design and methods*, Thousand Oaks, Calif., Sage Publications, 2003.

Articles

- « Avant-propos. Fondements, allocation et impact de l'aide », 2005, 19-2 *Revue d'économie du développement* p. 5-6.
- « The World Bank is finally embracing science, Editorial », *The Lancet*, 2004, n° 364 p. 731-732.
- AKRICH M., CALLON M., LATOUR B., « A quoi tient le succès des innovations? 2: Le choix des porte-parole », *Gérer et comprendre, Annales des Mines*, 1988.
- _____ « A quoi tient le succès des innovations? 1: L'art de l'intéressement », *Gérer et comprendre, Annales des Mines*, 1988.
- ANADON M., GUILLEMETTE F., « La recherche qualitative est-elle nécessairement inductive? », *Recherches Qualitatives*, Hors Série-5, 2007, p.26-37.
- AVENIER M.-J., C. THOMAS, « Mixer Quali et Quanti pour quoi faire? Méthodologie sans épistémologie n'est que ruine de la réflexion! », *Cahier de Recherche CERAG* 2011.
- AVVISATI F., GURGAND M., GUYON N., MAURIN E., « Getting Parents Involved: A Field Experiment in Deprived Schools », *Review of Economic Studies*, 81-1, 2014, p. 57-83
- BANERJEE A., DUFLO E., « L'approche expérimentale en économie du développement », *Revue d'économie politique*, 119-5, 2009, p. 691-726.
- BARBIER J.-C., MATYJASIK N., « Évaluation des politiques publiques et quantification en France : des relations ambiguës et contradictoires entre disciplines », *Revue Française de Socio-Economie*, 1-5, 2010, p. 123-140.
- BARBIER J.-C., « Eléments pour une sociologie de l'évaluation des politiques publiques en France », *Revue Française des Affaires Sociales* 1-2, 2010, p. 27-49.
- BARBIER J.-C., « Evaluations "expérimentales": quelques leçons d'histoire », *Politiques sociales et familiales* 110, 2012, .19-31.
- BARON G., MONNIER E., « Une approche pluraliste et participative: Coproduire l'évaluation avec la société civile », *Informations sociales* 110., 2003.
- BEHAGHEL L., CREPON B., GURGAND M., KAMIONKA T., LEQUIEN L., RATHELOT R., ZAMORA P., « L'accompagnement personnalisé des demandeurs d'emploi: les enseignements de trois expériences contrôlées en France », *Revue française d'économie*, 28-1, 2013, p. 123-158.
- BENNETT C., HOWLETT M., « The lessons of learning: Reconciling theories of policy learning and policy change », *Policy Sciences*.25, 1992, p. 275-294.
- BERGERON H., « Eléments pour une sociologie de l'entrepreneur-frontière », *Revue française de sociologie*, 54-2, 2013, 263-302.

- BONGRAND P., LABORIER P., « L'entretien dans l'analyse des politiques publiques: un impensé méthodologique? », *Revue Française de Science Politique*, 55-1, 2005, p. 73-111.
- BURTLESS G., « The Case for Randomized Field Trials in Economic and Policy Research », *The Journal of Economic Perspectives*, 9-2, 1995, p. 63-84
- CALLON M., « Eléments pour une sociologie de la traduction, la domestication des coquilles Saint-Jacques et des marins-pêcheurs dans la baie de Saint-Brieuc », *l'Année sociologique*, 36, 1986, p. 169-208.
- CAMPBELL D. T., « Assessing the impact of planned social change », *Evaluation and Program Planning*, 2, 1979, 67-90.
- CAMPBELL D. T., « Experiments as arguments », in HOUSE E.R, MATHISONS S, PEARSOL J.A., PRESKILL H., *Evaluation studies review annual*, 7, 1982, p. 117-127.
- CAMPBELL D. T., « Reforms as Experiments », *American Psychologist*, 24-4, 1969 p. 409-429.
- CARD D., DELLAVIGNA S., MALMENDIER U., « The Role of Theory in Field Experiments », *Journal of Economic Perspectives*, 25-3, 2011, p. 39-62.
- CHANUT V., « Contes et mécomptes d'une connaissance ordinaire », *Pouvoirs locaux*, 11-57, 2003 p. 99-105.
- CHELIMSKY E., « A Clash of Cultures: Improving the “fit” Between Evaluative Independence and the Political Requirements of a Democratic Society », *American Journal of Evaluation*, 29-4, 2008, p.400-415.
- COLETTI P., RADAELLI C. M., « Economic Rationales, Learning, and Regulatory Policy Instruments », *Public Administration*, 91-4, 2013, p.1056-1070.
- COOK T., SCRIVEN M., « Contemporary Thinking About Causation in Evaluation: A dialogue With Tom Cook and Michael Scriven », *American Journal of Evaluation*, 31-1, 2010, p. 105-117.
- CORCUFF P., « Un OVNI dans le paysage français ? Eléments de réflexion sur l'évaluation des politiques publiques en France », *politix*, 6-24, 1993, p.190-209.
- CRÉPON B., DUFLO E., GURGAND M., RATHELOT R., ZAMORA P., « Do Labor Market Policies have Displacement Effects? Evidence from a Clustered Randomized Experiment », *The Quarterly Journal of Economics*, 128-2, 2013, p. 531-580.
- DAVIES J. S., « The Limits of Joined-up Government: Towards a Political Analysis », *Public Administration*, 87-1, 2009, p. 80-96.
- DAVIES P., « The State of Evidence-Based Policy Evaluation and its Role in Policy Formation », *National Institute Economic Review*, 219, 2012, p.41-52.
- DAVIES P., « What is Evidence-Based Education? », *British Journal of Educational Studies*, 47-2, 1999, p.108-121.
- DEATON A., « Instruments of development: Randomization in the tropic, and the search for the elusive keys to economic Development », *Mimeo Princeton University* 2009.
- _____ « Instruments, Randomization, and Learning about Development », *Journal of Economic Literature*, 48, 2010, p.424-455.
- DEHUE T., « Deception, Efficiency, and Random Groups: Psychology and the Gradual Origination of the Random Group Design », *Isis*, 88-4, 1997, p. 653-673.
- _____ « Establishing the experimenting society: The historical origin of social experimentation according to the randomized controlled design », *American Journal of Psychology*, 114-2, 2001, p. 283-302.
- DELBRIDGE R., EDWARDS T., « Challenging conventions: Roles and processes during non-isomorphic institutional change », *Human Relations*, 61-3, 2008, p. 299-325.

- DER KNAAP P. VAN, « Theory-based Evaluation and learning: Possibilities and Challenges », *Evaluation*, 10-1, 2004, p. 16-34.
- DESAGE F., GODARD J., « Désenchantement idéologique et réenchantement mythique des politiques locales », *Revue Française de Science Politique*, 55-4, 2005, p. 633-661.
- DEVAUX-SPATARAKIS A., GREGOT A., « Les défis de l'emploi de l'étude de cas en évaluation », *Politiques Sociales et Familiales*, 110, 2012, p. 33-43.
- DEVAUX-SPATARAKIS A., « L'évaluation "basée sur la théorie" entre rigueur scientifique et contexte politique », *Politiques et Management Public*, 31-1, 2014, p. 75-91.
- DIMAGGIO P., POWELL W., « The Iron Cage Revisited: Institutional Isomorphism and Collective Rationality in Organizational Fields », *American Sociological Review* 48, 1983, p.147-160.
- DODIER N., BARBOT J. « Le temps des tensions épistémiques: Le développement des essais thérapeutiques dans le cadre du sida », *Revue française de sociologie*, 41-1, 2000, p. 79-118.
- DOLOWITZ D., MARSH D., « Who Learns What from Whom : a Review of the Policy Transfer Literature », *Political Studies*, 44-2, 1996, p. 342-357.
- DONALDSON S. I., CHRISTIE C. A., « The 2004 Claremont Debate: Lipsey versus Scriven. Determining causality in program evaluation and applied research. Should experimental evidence be the gold standard? », *Journal of Multidisciplinary Evaluation*, 3, 2005, p. 60-77.
- DUFLO E., MONTIALOUX C., « 4. Repenser la lutte contre la pauvreté », *Regards croisés sur l'économie* 10, 2011, p.59-62.
- DUFLO E., « Evaluer l'impact des programmes d'aide au développement: le rôle des évaluations par assignation aléatoire », *Revue d'économie du développement*, 19-2, 2005, p. 185-226.
- DULONG D., « Quand l'économie devient politique. La conversion de la compétence économique en compétence politique sous la Ve République », *Politix*, 9-35, 1996, p. 109-130.
- DUNCAN T., « Sailors, Scurvy and Science », *Journal of the Royal Society of Medicine* 90, 1997.
- DUNLOP C. A., RADAELLI C. M., « Systematising Policy Learning: From Monolith to Dimensions », *Political Studies*, 61-3, 2013, p. 599-619.
- DURAN P., MONNIER E., « Le développement de l'évaluation en France. Nécessités techniques et exigences politiques », *Revue Française de Science Politique*, 42-2, 1992, p.235-262.
- EISENSTADT S., « Cultural orientations, institutional entrepreneurs, and social change: Comparative analysis of traditional civilizations », *American Journal of Sociology*, 85, 1980, p.840-869.
- EVANS M., DAVIES J., « Understanding policy transfer: a multi-level, multi-disciplinary perspective », *Public Administration*, 1999.
- FAURE A., SMITH A., « L'évaluation objet de savoir et enjeu de pouvoir, le cas des politiques publiques locales », *Sciences de la Société*, 32, 1994, p.101-111.
- FERRACCI M., MARTIN F., « Evaluation par expérimentation aléatoire de dispositifs de reclassement innovants pour les demandeurs d'emploi », *Travail et emploi*, 3-135, 2013, p. 5-26.
- FLIGSTEIN N., MCADAM D., « Toward a General Theory of Strategic Action Fields », *Sociological Theory*, 29-1, 2011, p. 1-26.
- FOUGERE D., « Expérimenter pour évaluer les politiques d'aide à l'emploi: les exemples anglo-saxons et nord-européens. », *Revue Française des Affaires Sociales*, 2000.

- GARUD R., JAIN S., KUMARASWAMY A., « Institutional entrepreneurship in the sponsorship of common technological standards: The case of Sun Microsystems and Java. », *Academy of Management Journal*, 45-1, 2002, p. 196-214.
- GERRING J., « What Is a Case Study and What Is It Good for? », *The American Political Science Review*, 98-2, 2004, p. 341-354.
- HACKING I., « Telepathy: Origins of Randomization in Experimental Design », *Isis*, 79-3, 1988, p. 427-451
- HECKMAN J. J., SMITH J. A., « Assessing the Case for Social Experiments », *The journal of Economic perspectives*, 9-2, 1995, p. 85-110.
- HOPPE R., « Policy analysis, science, and politics: from “speaking truth to power” to “making sense together” », *Science and Public Policy*, 26-3, 1999, p. 201-210.
- HOROWITZ I. L., « Social science mandarins: Policymaking as a political formula », *Policy Science*, 1-3, 1970, p. 339-360.
- KNAAP L. VAN DER, LEEUW F., BOGAERTS S., NIJSSEN L., « Combining Campbell Standards and the Realist Evaluation Approach, The Best of Two Worlds? », *American Journal of Evaluation*, 29-48, 2008, p. 48-57.
- L'HORTY Y., « Paradoxe de l'évaluation du RSA », *Projet*, 1-308, 2009, p. 21-29.
- L'HORTY Y., DUGUET E., PETIT P., « L'accompagnement des collégiens dans leur recherche de stage: une évaluation expérimentale », *Education & formation* 81, 2012.
- L'HORTY Y., DUGUET E., PETIT P., « Une évaluation expérimentale d'un micro-programme social », *Revue française d'économie*, 27-1, 2012, p. 107-127.
- LABROUSSE A., « Eléments pour un institutionnalisme méthodologique : autonomie, variation d'échelle, réflexivité et abduction », *Economie et Institutions*, 8, 2006, p.5-53.
- « Nouvelle économie du développement et essais cliniques randomisés: une mise en perspective d'un outil de preuve et de gouvernement », *Revue de la régulation*, 7, 2010.
- LALONDE R. J., « Evaluating the Econometric Evaluations of Training Programs with Experimental Data », *The American Economic Review*, 76-4, 1986, p. 604-620.
- LASCOUMES P., SIMARD L., « L'action publique au prisme de ses instruments », *Revue française de science politique*, 61-1, 2011, p. 5-22.
- LAURENT C., « Pourquoi s'intéresser à la notion d'“evidence-based policy”? », *Revue Tiers Monde*, 200, 2009, p. 853-873.
- LECA J., « L'évaluation dans la modernisation de l'État », *Politiques et Management Public*. 1993, p. 11-2
- LE GALÈS P., LASCOUMES P., « Introduction : Understanding Public Policy through its Instruments - From the Nature of Instruments to the Sociology of Public Policy Instrumentation », *Governance: An International Journal of Policy, Administration, and Institutions*, 20-1, 2007, p.1-21.
- LEVY D., SCULLY M., « The Institutional Entrepreneur as Modern Prince: The Strategic Face of Power in Contested Fields », *Organization Studies*, 28-7, 2007, p. 971-991.
- LEWIN A., MAURIN E., « Quels effets attendre de réformes permettant de cumuler emploi et aide sociale? Une évaluation sur données expérimentales américaines », *Revue économique*, 53-6, 2002, p. 1357-1368.
- LINDBLOM C., « The Science of “Muddling Through” », *Public Administration Review*, 19, 1959, p. 79-88.
- MAGUIRE S. HARDY C., « Discourse and Deinstitutionalization: The Decline of DDT », *Academy of Management Journal*, 52-1, 2009, p. 148-178.

- MARSTON G., WATTS R., « Tampering with the Evidence: A Critic Appraisal of Evidence-Based Policy-Making », *The Drawing Board: An Australian Review of Public Affairs*, 3-3, 2003, p. 143-163.
- MAYNE J., « Addressing Attribution Through Contribution Analysis: Using Performance Measures Sensibly », *The Canadian Journal of Program Evaluation*, 16-1, 2001, p. 1-24.
- MEHAIGNERIE P., « Décentralisation, expérimentation et évaluation », *Pouvoirs locaux*, 11-57, 2003, p.113-116.
- MONNIER E., HENARD F., « L'Evaluation des politiques publiques dans les collectivités territoriales : exercices imposés et initiatives propres », *Pouvoirs locaux*, 2000.
- NIOCHE J.-P., « De l'évaluation à l'analyse des politiques publiques », *Revue Française de Science Politique*, 32-1, 1982, p. 32-61.
- OLIVIER C., « strategic responses to institutional processes », *Academy of Management Review*, 16, 1991, p. 145-179.
- PARIENTE W., « Analyse d'impact : l'apport des évaluations aléatoires », *Stateco* 103, 2008.
- PARSONS W., « From Muddling Through to Muddling Up - Evidence Based Policy Making and the Modernisation of British Government », *Public Policy and Administration*, 17-3, 2002, p. 43-60.
- PETROSINO A., « Reflections on the Genesis of the Campbell Collaboration », *The Experimental Criminologist*, 8, 2013, p. 9-12.
- PINSON G., SALA PALA V., « L'entretien en sociologie de l'action publique », *Revue Française de Science Politique*, 57-5, 2007, p. 555-597.
- RAVALLION M., « Should Randomistas rule? », *Economist's Voice*, 2009.
- RIEPER H., « The evidence movement, the development and consequences of methodologies in review practices », *Evaluation*, 15, 2009, p. 141-163.
- ROBERTSON D., « Political Conflict and Lesson Drawing », *Journal of Public Policy*, 11-1, 1991, p. 55-91.
- ROSE R., « What is Lesson-drawing? », *Journal of Public Policy*, 11, 1991.
- ROSSI P. H., WRIGHT S. . « Evaluation research: An assessment of theory, practice, and politics », *Evaluation Quarterly*, 1, 1977, p. 5-52.
- RUBIN D. B., « Estimating causal effects of treatments in randomized and non-randomized studies », *Journal of Educational Psychology*, 66-5, 1974, p. 688-701.
- SHADISH W. R., « Evaluation Theory is Who We Are », *American Journal of Evaluation*, 19-1, 1998, p.1-19.
- SHAPIN S., « The House of Experiment in Seventeenth-Century England », *Isis*, 79-3, 1988, p. 373-404.
- BEN SLIMANE K., LECA B., « Le travail institutionnel : origines théoriques, défis et perspectives », *Management & Avenir*, 7-37, 2010, p. 53-69.
- SMITH N. L., « The feasibility and desirability of experimental methods in evaluation », *Evaluation and Program Planning*, 3, 1981.
- SOULE B., « Observation participante ou participation observante? Usages et justifications de la notion de participation observante en sciences sociales », *Recherches Qualitatives*, 27-1, 2007, p. 127-140.
- SPENLEHAUER V., « Intelligence gouvernementale et sciences sociales », *Politix*, 12-48, 1999, p. 95-128.
- SUCHMAN M. C., « Managing Legitimacy: Strategic and Institutional Approaches », *Academy of Management Review*, 20, 1995, p. 571-610.

- TSANG E. W. K., KWAN K.-M., « Replication and Theory Development in Organizational Science: A Critical Realist Perspective », *The Academy of Management Review*, 24-4, 1999, p. 759-780,
- VARONE F., JACOB S., « Cheminement institutionnel de l'évaluation des politiques publiques en France, en Suisse et aux Pays-Bas (1970-2003) », *pomap*, 22-2, 2004, p. 135-152.
- VARONE F., JACOB S., « Institutionnalisation de l'évaluation et Nouvelle Gestion Publique: un état des lieux comparatif », *Revue Internationale de Politique Comparée*, 11-2, 2004, p. 271-292.
- VARONE F., « Le choix des instruments de l'action publique: analyse comparée des politiques énergétiques en Europe et Amérique du nord », *Revue Internationale de Politique Comparée*, 7-1, 2000, p. 167-201.
- VEDUNG E., « Four waves of evaluation diffusion », *Evaluation*, 3-16, 2010, p. 263-277.
- WEISS C. H., « The Fairy Godmother and Her Warts: Making the Dream of Evidence-Based Policy Come True », *American Journal of Evaluation*, 29-29, 2008.
- _____ « Where Politics and Evaluation Research Meet », *Evaluation Practice*, 14-1, 1993, p. 93-106.
- ZAMORA P., « La méthode d'évaluation aléatoire: apports et limites », *Tracés*, 2011, p.175-186.

Thèses, Mémoires

- LEQUIEN L., *Essais d'évaluation de politique publique dans les champs de l'éducation, de la santé et des politiques d'emploi*, Paris, Ecole des Hautes Etudes en Sciences Sociales, 2011.
- MATYJASIK N., *L'évaluation des politiques publiques dans une France décentralisée, institutions, marché et professionnels*, Université Montesquieu-Bordeaux IV/Sciences-po Bordeaux, 2010.
- RIEGERT A., *Rapport de stage effectué auprès du Haut-commissaire à la jeunesse du 4 mai 2009 au 21 août 2009*, Paris School of Economics, Ecole normale supérieure, Ecole des Hautes Etudes en Sciences Sociales, 2009.
- SPENLEHAUER V., *L'évaluation des politiques publiques, avatar de la planification*, Grenoble, Université Pierre Mendès-France, 1998.

Littérature grise, rapports, documents de travail

- « Programme », dans Conférence nationale de l'expérimentation sociale « Réformes contre placebo? », Théâtre National de Chaillot-Paris, 2010.
- ABDUL LATIF JAMEEL POVERTY ACTION LAB, « Accompagnement renforcé des bénéficiaires du RMI, dans les Hauts de Seine, France », Abdul Latif Jameel Poverty Action Lab, en ligne : <<http://www.povertyactionlab.org/fr/evaluation/accompagnement-renforc%C3%A9-des-b%C3%A9n%C3%A9ficiaires-du-rmi-dans-les-hauts-de-seine-france>> (consulté le 1 août 2013).
- ABDUL LATIF JAMEEL POVERTY ACTION LAB. « Getting parents involved », *J-PAL Briefcase*, février 2011.
- ACTEURS PUBLICS, « Le cabinet de Martin Hirsch », acteurspublics.com (23 juillet 2008), en ligne : <<http://www.actorspublics.com/2008/07/23/le-cabinet-de-martin-hirsch>> (consulté le 29 juillet 2013).
- AFIJ, « Rebond », Note de restitution finale du porteur de projet, Mission d'animation du Fonds d'expérimentations pour la jeunesse, 2013.
- AGENCE NOUVELLE DES SOLIDARITES ACTIVES, « Revue succincte d'activité », 2, 2006.

- AGENCE NOUVELLE DES SOLIDARITES ACTIVES. « Séance 2- L'expérimentation sociale: innover par la preuve? » Programme, 25 novembre 2010.
- AGENCE NOUVELLE DES SOLIDARITES ACTIVES, « Revue succincte d'activité », 1, 2006.
- AGENCE NOUVELLE DES SOLIDARITES ACTIVES, « Rapport d'activités », année 2006, 2007.
- ALDEGHI I., JAUNEAU P., MULLER L., « Bilan méthodologique des évaluations conduites par le Crédoc dans le cadre du Fonds d'expérimentation pour la jeunesse », 300, coll. *Cahier de recherche*, CREDOC, 2012.
- ALLEGRE G. « L'expérimentation sociale des incitations financières à l'emploi, questions méthodologiques et leçons des expériences nord-américaines », *Document de travail de l'OFCE*, n°2008-22, 2008.
- ALTON-LEE A., « A Collaborative Knowledge Building Strategy to Improve Educational Policy and Practice: Work-in-Progress in the Ministry of Education's Iterative Best Evidence Synthesis Programme », 2004.
- ALTON-LEE A., « Iterative Best Evidence Synthesis: Strengthening Research, Policy and Practice Links to Improve Outcomes », 2006.
- ARTUS P. (dir.), « Réformer l'Enseignement supérieur et la Recherche, un pacte pour une nouvelle université », *Le Cercle des économistes*, 5.135, 2004.
- ASSOCIATION NATIONALE DES DIRECTEURS DE MISSION LOCALE, « Quels enseignements retenir des expérimentations sociales sur l'alternance menées par les Missions Locales? », 2013.
- BANKS G., « Challenges of Evidence- Based Policy- Making, Australian Government, Productivity Commission », Australian Public Service Commission, 2009.
- BEFANI B., « Models of Causality and Causal Inference », in *Broadening the Range of Designs and Methods for Impact Evaluation*, coll. DFID, 2012.
- BEHAGHEL L. GURGAND M., « Programme expérimental « Bourse aux projets de classe »: bilan de la phase pilote du point de vue de l'évaluateur », 2010.
- BEHAGHEL L., CRÉPON B., GURGAND M., « Private and Public Provision of Counseling to Job-Seekers: Evidence from a Large Controlled Experiment », IZA DP No. 6518, coll. *Discussion paper series*, Bonn Germany, IZA, 2012.
- BEHAGHEL L., CREPON B., LE BARBANCHON T., « Evaluation de l'impact du CV anonyme, synthèse et rapport final », PSE, CREST, J-Pal, 2011.
- BEHAGHEL L., CRÉPON B., GURGAND M., LE BARBANCHON T., « Please Call Again: Correcting Non-Response Bias in Treatment Effect Models », IZA DP No. 6751, coll. *Discussion Paper series*, Bonn Germany, IZA, 2012.
- BEHAGHEL L., CRÉPON B., GURGAND M., LE BARBANCHON T., « Sample attrition bias in randomized experiments », halshs.archives-ouvertes, février 2011.
- BEHAGHEL L., DE CHAISEMARTIN C., CHARPENTIER A., GURGAND M., « Internats d'excellence: les enseignements de Sourdun », J-PAL Europe, Institut des Politiques Publiques, avril 2013.
- BEHAGHEL L., CHIODI V., GURGAND M., « Evaluation de l'impact du programme de parrainage d'aide à l'orientation de l'association Actenses », Rapport final, Rapport d'évaluation, Institut des Politiques Publiques, J-PAL Europe, 2013.
- BLUNKETT D. « Blunkett rejects anti-intellectualism and welcomes sound ideas », DfEE News, 2000.
- CABINET OFFICE, « Adding it up, improving analysis and modelling in Central Government », London, Performance and Innovation Unit, 2000.
- CABINET OFFICE, « Professional Policy Making for the Twenty-First Century », London, Cabinet Office, 1999.
- CABINET OFFICE, « Modernising Government White Paper », London, Center for Management and Policy Studies, 1999.

- CALVES G., « Histoire et perspectives du cadre juridique des expérimentations législatives et réglementaires en France », Paris, Dares, 2008, p. 95–106.
- CHARPAIL C., KLEIN T., ZILBERMAN S., « Evaluation des politiques d'emploi: la deuxième génération des panels des bénéficiaires », n°95, coll. *Document d'études*, Dares, 2005.
- COALITION FOR EVIDENCE-BASED POLICY, « Bringing Evidence-Driven Progress To Crime and Substance-Abuse Policy: A Recommended Federal Strategy », 2003.
- COALITION FOR EVIDENCE-BASED POLICY, « Hierarchy of Study Designs for Evaluating the Effectiveness of STEM Education Project or Practice », 2007.
- COALITION FOR EVIDENCE-BASED POLICY, « Bringing Evidence-Driven Progress To Education: A recommended Strategy for the U.S. Department of Education », 2002.
- COMETS, « Avis du COMETS sur « l'éthique de la recherche dans l'expérimentation sociale » », 19 janvier 2010.
- COMITE D'EVALUATION DES EXPERIMENTATIONS, « Rapport d'étape sur l'évaluation des expérimentations RSA », Paris, 2008.
- COMITE D'EVALUATION DES EXPERIMENTATIONS, « Rapport final sur l'évaluation des expérimentations RSA », Paris, 2009.
- COMITE INTERMINISTERIEL DE LA JEUNESSE, « Priorité Jeunesse », 2013.
- COMMISSION FAMILLES, VULNERABILITE, PAUVRETE, « Au possible, nous sommes tenus. La nouvelle équation sociale, 15 résolutions pour combattre la pauvreté des enfants », 2005.
- CONSEIL SCIENTIFIQUE DU FEJ, « Guide méthodologique pour l'évaluation des expérimentations sociales à l'intention des porteurs de projets », 2009.
- CONSEIL SCIENTIFIQUE DU FEJ, « Rapport du Conseil Scientifique du Fonds d'expérimentation pour la jeunesse pour l'année 2011 », Fonds d'expérimentation pour la jeunesse, 2012.
- CONSEIL SCIENTIFIQUE DU FEJ, « Rapport du Conseil scientifique du Fonds d'expérimentations pour la jeunesse, pour la période mai 2009-décembre 2010 », Ministère de l'éducation nationale de la jeunesse et de la vie associative, 2011.
- COUR DES COMPTES, « Rapport public annuel, Des internats d'excellence à ceux de la réussite: la conduite chaotique d'une politique éducative et sociale », février 2014.
- CREDOC, « AgiLoJe, Rapport d'évaluation, Fonds d'expérimentations pour la jeunesse », 2012.
- CREDOC, « Note d'étape sur l'évaluation « Un pas dehors » », Mission d'animation du Fonds d'expérimentations pour la jeunesse, 31 décembre 2010.
- CREPON B., DUFLO E., GURGAND M., RATHELOT R., ZAMORA P., « L'accompagnement des jeunes diplômés demandeurs d'emploi par des opérateurs privés de placement: les enseignements d'une évaluation », *Dares Analyses*, 94, 2011.
- CRÉPON B., DEJEMEPPE M., GURGAND M., « Counseling the Unemployed: Does It Lower Unemployment Duration and Recurrence? », *discussion paper series* IZA 1796, 2005.
- CRÉPON B., M. FERRACCI et D. FOUGÈRE, « Training the Unemployed in France: How Does it Affect Unemployment Duration and Recurrence? », coll. *Discussion paper series*, Bonn Germany, IZA, 2007.
- CRÉPON B., GURGAND M., KAMIONKA T., LEQUIEN L., « Is Counseling Welfare Recipients Cost-Effective? Lessons from a Random Experiment »
- CREST, « Note d'étape sur l'évaluation, Rebond, pour les étudiants décrocheurs de l'enseignement supérieur », Mission d'animation du Fonds d'expérimentations pour la jeunesse, 31 décembre 2010.
- CREST, « Rebond: programme d'accompagnement des étudiants décrocheurs de l'enseignement supérieur », Rapport d'évaluation, Fonds d'expérimentation pour la jeunesse, 2012.
- CREST-DARES, « Sécurisation du parcours des jeunes s'engageant dans l'apprentissage : bilan d'une expérimentation aléatoire contrôlée conduite en Corrèze », Rapport

- d'évaluation, Ministère des Sports, de la Jeunesse, de l'Education populaire et de la Vie associative, Fonds d'expérimentation pour la Jeunesse, 2012.
- DARES-POLE EMPLOI, « L'accompagnement renforcé des demandeurs d'emploi: évaluation du recours aux opérateurs privés par Pôle emploi de 2009 à 2011 », 15, coll. *Les cahiers études*, Direction des études et de l'évaluation, 2013.
- DELARUE J., NAUDET J. D., SAUVAT V., « Les évaluations sont-elles utiles? Revue de la littérature sur « connaissances et décisions » », 03, coll. *Série Notes méthodologiques*, Ex Post, Département de la Recherche, Division Evaluation et Capitalisation, Agence Française de Développement, 2009.
- DELEAU M., « Evaluer les politiques publiques: méthodes, déontologie, organisation: rapport du Groupe de travail « Méthodes d'évaluation des politiques publiques » », coll. Plan/Commissariat général du plan, n°13, Paris, Documentation française, 1986.
- DEMUINCK C., « Le décrochage universitaire », Ministère de l'Enseignement supérieur et de la Recherche, 2011.
- DUFLO E. « Chaire Savoirs contre pauvreté », La lettre du Collège de France, 25.5, 2009.
- DUFLO E., GLENNISTER R., KREMER M., « Randomized Evaluations of Interventions in Social Service Delivery », MIT. no date
- DU PARQUET L., STEUNOU P.-Y., « « Améliorer la Qualité des Emplois Exercés par les Etudiants » (AQ3E) », note de restitution finale.
- ECOLE D'ECONOMIE DE PARIS , « Les effets de l'internat d'excellence de Sourdon sur les élèves bénéficiaires : résultats d'une expérience contrôlée », Rapport d'évaluation, Fonds d'expérimentation pour la jeunesse, 2013.
- ECOLE D'ECONOMIE DE PARIS , « Quels effets attendre d'une politique d'implication des parents dans les collèges, évaluation de l'impact de la Mallette des parents, Fonds d'expérimentations pour la jeunesse », 2011.
- ECOLE D'ECONOMIE DE PARIS. « Note d'étape sur l'évaluation des Internats d'Excellence », Mission d'animation du Fonds d'expérimentations pour la jeunesse, 30 juin 2010.
- EUROPEAN EVALUATION SOCIETY. « EES Statement: The importance of a methodologically diverse approach to impact evaluation-specifically with respect to development aid and development interventions », décembre 2007.
- EVALUATION GAP WORKING GROUP, « When Will We Ever Learn? Improving Lives Through Impact Evaluation », Center for Global Development, 2006.
- FEDERATION TRAVAIL EMPLOI ET POLITIQUES PUBLIQUES, « Tous en stage », Rapport d'évaluation, Fonds d'expérimentation pour la jeunesse, 2012.
- FONDS D'EXPERIMENTATION POUR LA JEUNESSE, « Premiers enseignements des expérimentations, synthèses thématiques 2013 », Ministère des Sports, de la Jeunesse, de l'Education populaire et de la Vie associative, 2013.
- , « Rapport d'activité 2009-2011 », 2012.
- , « Rapport d'activité 2009-2011, Annexes », 2012.
- , « Rapport d'activité 2012 », Ministère des sports, de la jeunesse, de l'éducation populaire et de la vie associative, 2013.
- , « Programme national d'expérimentations pour les jeunes, premier appel à projets 6 avril 2009 », 7 mai 2009.
- , « Appel à projet n°4, « Favoriser le parcours des jeunes vers l'autonomie par les pratiques artistiques et culturelles » », 8 juillet 2010.
- GERTLER P., « The Impact of Progresa on Health, International Food Policy Research Institute », 2000.
- GOMEL B., SERVERIN E., « Expérimenter pour décider? Le RSA en débat », *Document de travail*, Centre d'étude de l'Emploi 2009.119.
- HASKINS R., BARON J., « The Obama Administration's evidence-based social policy initiatives: An overview », Coalition for Evidence-Based Policy. no date
- HAUT COMMISSARIAT A LA JEUNESSE, « La mission Valletoux-Noblecourt », 2009.

- HAUT COMMISSARIAT AUX SOLIDARITES ACTIVES CONTRE LA PAUVRETE , « Rencontres de l'expérimentation sociale en Europe, Programme », 21 novembre 2008.
- HERCK M., « The coalition for evidence-based policy: its role in advancing evidence-based reform, 2004-2009 », William T. Grant Foundation, 2009.
- HIGGINS J., GREEN S., (dir.). « Cochrane Handbook for Systematic Reviews of Interventions », Version 5.1.0, The Cochrane Collaboration, 2011.
- INDEPENDANT EVALUATION GROUP, « World Bank Group Impact Evaluations: Relevance and Effectiveness », Washington, D.C, World Bank, 2012.
- IREDU/CNRS, « Inscrire les contrats en alternance dans une logique de parcours sécurisé », Rapport d'évaluation, Ministère de l'Education nationale, de la Jeunesse et de la Vie associative, 2012.
- JOUE M., « Les expérimentations sociales en Europe, Vers une palette plus complète et efficace de l'action communautaire en faveur de l'innovation sociale », Notre Europe, 2008.
- « Journée d'étude de la S.F.E. L'évaluation et les activités voisines: de la concurrence à la complémentarité », Programme, 11 décembre 2007.
- JOWELL R., « Trying it out : the role of « pilots » in policy-making : report of a review of government pilots », Great Britain. Cabinet Office. Strategy Unit, London, Cabinet Office, Strategy Unit, 2003.
- J-PAL EUROPE, « Social Experimentation, A methodological guide for policy makers », Directorate General for Employment, Social Affairs and Inclusion, European Commission, 2011.
- J-PAL, « Accompagnement renforcé des bénéficiaires du RMI, dans les Hauts de Seine, France », Poverty Action Lab, en ligne : <http://www.povertyactionlab.org/fr/evaluation/accompagnement-renforc%C3%A9-des-b%C3%A9n%C3%A9ficiaires-du-rmi-dans-les-hauts-de-seine-france> (consulté le 13 novembre 2013).
- L'HORTY Y., PETIT P., « L'évaluation aléatoire: un succès qui ne doit rien au hasard », la vie des idées (2011).
- _____, « Evaluation aléatoire et expérimentations sociales », 135, coll. *Document de travail*, Centre d'études de l'emploi, 2010.
- LABORATOIRE ERUDITE, « Améliorer la Qualité des Emplois Exercés par les Etudiants (AQ3E) », Rapport d'évaluation, 2013.
- LACOUETTE-FOUGERE C., LASCOUMES P., « Les scènes multiples de l'évaluation, les problèmes récurrents de son institutionnalisation », Laboratoire interdisciplinaire d'évaluation des politiques publiques, Sciences-po Paris, 2003.
- « Lancement du Grenelle de l'insertion, les rencontres de l'expérimentation Sociale, programme », 2007.
- LE HAUT COMMISSAIRE AUX SOLIDARITES ACTIVES CONTRE LA PAUVRETE, « Communiqué de presse, Bilan des Rencontres de l'expérimentation sociale en Europe », Présidence française du Conseil de l'Union européenne 2008, 22 novembre 2008.
- _____, « Appel à projet d'expérimentation sociale 2008 », 2007.
- LEE J., « Is Evidence-Based Government Possible? », 2004.
- LES MEMBRES DU CONSEIL D'ANALYSE ECONOMIQUE, « Evaluation des politiques publiques », *Les notes du conseil d'analyse économique*, n°1, Conseil d'analyse économique, 2013.
- « Lettre de mission de M. Nicolas SARKOZY », Président de la République, adressée à M. Martin Hirsch, Haut commissaire auprès du Premier ministre, aux Solidarités actives contre la pauvreté, 2007.
- LEVITT S., LIST J., « Field Experiments in Economics: the Past, the Present, and the Future », National Bureau of Economic Research, Cambridge MA, 2008.
- MAFEJ, « AP1 Statistiques sur les projets déposés et les projets retenus », HCJ, DJEPVA, 15 mars 2010.

- MAFEJ, « Programme national d'expérimentations pour la jeunesse, deuxième appel à projets », 9 octobre 2009.
- MAILLARD D., MORA V., ROUAUD P., THERY M., VENEAU P., LEHBERGER T., « Evaluation d'un dispositif d'accompagnement de jeunes en rupture scolaire », Net.Doc 2012.97.
- MARTIN B., « The coalition for evidence-based policy: its impact on policy and practice », William T. Grant Foundation, 2004.
- MISSION LOCALE DE BEAUNE, « Inscrire les contrats en alternance dans une logique de parcours sécurisé », Note de restitution finale, Fonds d'expérimentation pour la jeunesse, 2012.
- _____, « Note d'étape sur l'expérimentation, Inscrire les contrats en alternance dans une logique de parcours sécurisé », Mission d'animation du Fonds d'expérimentations pour la jeunesse, 31 décembre 2010.
- NAUDET J. D., DELARUE J., « Fostering Impact Evaluations at Agence Française de Développement: A Process of In-house Appropriation and Capacity-Building », Network of Networks on Impact Evaluation, janvier 2008.
- _____, « Promouvoir les évaluations d'impact à l'Agence Française de Développement : renforcer l'appropriation et l'apprentissage institutionnels », 02, coll. *Série Notes méthodologiques Ex Post*, Département de la Recherche, Division Evaluation et capitalisation, Agence Française de Développement, 2007.
- NOBLECOURT O., VALLETOUX F., « L'expérimentation sociale locale en France: bilan, enjeux, perspectives », Rapport de mission 2009-2010, rapport non publié.
- O'DWYER L., « A critical review of evidence-based policy making », 58, Australian Housing and Urban Research, 2004.
- PARIS SCHOOL OF ECONOMICS, « La lettre: spéciale 5 ans », n°8, janvier 2012.
- _____, « Rapport d'activité 2010-2011, la science économique au service de la société », 2012.
- PAWSON R., « Assessing the quality of evidence in evidence-based policy: why, how and when? », ESRC Research Methods Programme, working paper.
- PAWSON R., TILLEY N., « Realist Evaluation », British Cabinet Office, 2004.
- PETIT P., DUGUET E., L'HORTY Y., « Discrimination résidentielle et origine ethnique: Une étude expérimentale en Ile-de-France », TEPP Rapport de recherche, octobre 2012.
- PREFECTURE DE REGION RHONE ALPES. « Appel à pistes d'innovation et à expérimentation sociales 2007, financé sur le Fonds d'innovation et d'Expérimentation Sociales de la DIIIESES », 2007.
- PROGRAM EVALUATION AND METHODOLOGY DIVISION, « Case Study Evaluations », United States General Accounting Office, 1990.
- RECTORAT DE CRETEIL, « La Mallette des parents, APDIIIESES-11, Note de restitution finale remise par le Rectorat de Créteil, porteur du projet, au Fonds d'Expérimentations pour la Jeunesse dans le cadre de l'appel à projets lancé en 2008 par le Ministère en charge de la Jeunesse », Rapport d'évaluation, 2011.
- REUSSIR MOI AUSSI, « « Tous en Stage », Note de restitution finale du porteur de projet », Fonds d'expérimentation pour la jeunesse, 2011.
- ROMANI C., « Le Céreq et l'évaluation des expérimentations sociales », 84, coll. Net.Doc, Céreq, 2011.
- ROSS H. L., « Testing Experimentation: Income Maintenance and Social Policy », Cambridge, Mass, Edited by James G. Abert and Murray Kamrass.Ballinger, 1974.
- SARKOZY N., « Agir pour la jeunesse, Discours de M. Le Président de la République Française », Avignon, 2009.
- SIEBEL C., AVENTURE F., LARDY S., « L'évaluation de l'accompagnement renforcé des demandeurs d'emploi », 2010.
- SOCIETE FRANÇAISE D'EVALUATION. « L'évaluation des impacts des programmes et services publics », Cahier de la SFE n°6, septembre 2011.

- SOLESBURY W., « Evidence Based Policy: Whence it Came and Where it's Going », ESRC UK Centre for Evidence Based Policy and Practice, Queen Mary, University of London, octobre 2001.
- TEPP-CNRS, « Faut-il subventionner le permis de conduire des jeunes en difficulté d'insertion? 10 000 permis pour réussir », Rapport d'évaluation quantitative, Fonds d'expérimentation pour la jeunesse, 2012.
- UNION EUROPEENNE. DIRECTORATE-GENERAL FOR REGIONAL POLICY AND COHESION, « Evaluer les programmes socio-économiques : Principales techniques et outils d'évaluation », 3, coll. MEANS, Luxembourg, Office des publications officielles des Communautés européennes, 1999.
- UNIVERSITE DU MAINE. « Note d'étape sur l'expérimentation AQ3E », Mission d'animation du Fonds d'expérimentations pour la jeunesse, 31 décembre 2010.
- VAESSEN J., « Challenges in impact evaluation of development interventions: opportunities and limitations for randomized experiments », University of Antwerp 2010.
- WHITE H. « Theory-Based Impact Evaluation: Principles and Practice, International Initiative for Impact Evaluation », 2009.

Articles de presse, sites internet

- « Eric Zemmour : "Les CV anonymes reposent sur des présupposés d'une France raciste" », RTL.fr (13 avril 2011), en ligne : <<http://www.rtl.fr/voir/eric-zemmour-les-cv-anonymes-reposent-sur-des-presupposes-d-une-france-raciste-7676862560>> (consulté le 31 mars 2014).
- « Jean-Benoît Dujol, Un jeune au service des jeunes », *Les Echos*, 3 juin 2010, p. 11.
- « President Barack Obama's Inaugural Address | The White House », en ligne : <http://www.whitehouse.gov/the_press_office/President_Barack_Obamas_Inaugural_Address> (consulté le 2 mai 2014).
- ARIDJ J., « L'accompagnement renforcé plus efficace pour les chômeurs », *Le point*, 4 juillet 2008.
- ARMANET F., « Comment lutter contre la pauvreté: le face-à-face Esther Duflo-Martin Hirsch », *Le Nouvel Observateur* n°2359, 21 janvier 2010, p. 70-73.
- ASSOCIATION DES JOURNALISTES DE L'INFORMATION SOCIALE, « Martin HIRSCH - Président d'Emmaüs France et de l'Agence nouvelle des solidarités actives - Petit-déjeuner » (20 juin 2006), en ligne : <<http://www.ajis.asso.fr/ajis/content/martin-hirsch-president-demmaus>> (consulté le 25 juillet 2013).
- BAUMARD M., « L'internat d'excellence fait progresser les élèves en maths, pas en français », *Le Monde*, 11 avril 2013.
- BEYER C., « Vers la fin des internats d'excellence », *Le Figaro*, 11 avril 2013.
- BOURGUIGNON F., « Le RSA montre la voie de l'évaluation », *Les Echos*, 3 juin 2009.
- _____ « PSE-Ecole d'économie de Paris : après 5 ans d'existence, un nouveau chapitre s'ouvre... », Communiqué de presse, 16 février 2012.
- BOUVET L., « Forum "la nouvelle critique sociale" - Grenoble 12-14 mai 2006 », L(B)LOG (30 mai 2006), en ligne : <<http://laurentbouvet.wordpress.com/2006/05/30/forum-la-nouvelle-critique-sociale-grenoble-12-14-mai/>> (consulté le 26 juillet 2013).
- DUFLO E., « L'économie du développement à l'épreuve du terrain », *La vie des idées.fr*, 5 mai 2009.
- DUFLO E., « Esther Duflo, pratiquer l'économie comme une vraie science humaine », *Madame Figaro*, 28 février 2009.

- DUJOL J.-B., « Incentives matter », *Le tigre*, 6, octobre 2007, p. 14-15.
- HIRSCH M., « Martin Hirsch : « Je ne suis ni dupe ni naïf » », *Le monde*, 20 mai 2007, p. 6.
- , « Martin Hirsch “Il y avait un climat dans lequel je ne me sentais pas à l’aise” », *Le monde*, 23 mars 2010.
- , « Au possible nous sommes tenus », dossier de presse, Haut Commissaire aux solidarités actives, 20 septembre 2007.
- , « Contre la misère, osons ! », *Le Monde*, 2 février 2006.
- KAHN A., MALINGRE V., « Le Monde.fr : Les « French economists » font école », *Le Monde*, 22 février 2007.
- LA REPUBLIQUE DES IDEES, « La nouvelle critique sociale, Grenoble, du 12 au 14 Mai 2006 », La République des idées, en ligne : <<http://www.repid.com/La-nouvelle-critique-sociale,7.html>> (consulté le 26 juillet 2013).
- LAMARRE M., « Fonds d’expérimentation Jeunesse : ou comment l’Etat s’auto-subsventionne », Paperblog, en ligne : <<http://www.paperblog.fr/2129739/fonds-d-experimentation-jeunesse-ou-comment-l-etat-s-auto-subsventionne/>> (consulté le 5 novembre 2013).
- LANDRE M., « Les réseaux de Martin Hirsch », L’Expansion.com (1 mai 2006), en ligne : <http://lexpansion.lexpress.fr/economie/les-reseaux-de-martin-hirsch_25457.html> (consulté le 29 juillet 2013).
- LECLERE T., « Ça s’appelle “La République des idées” », *Télérama* n°3047, 7 juin 2008.
- LEFEBVRE E., BOURGUIGNON F., HIRSCH M., « La crise rend la réforme plus indispensable que jamais », *Les Echos*, 22 mai 2009.
- LEVY E., « Miracle chez Hirsch: la Résurrection de Saint Martin », Marianne, en ligne : <http://www.marianne.net/Miracle-chez-Hirsch-la-Resurrection-de-Saint-Martin_a198191.html> (consulté le 31 octobre 2013).
- ORSZAG P., « Building Rigorous Evidence to Drive Policy | The White House » (8 juin 2009), en ligne : <<http://www.whitehouse.gov/omb/blog/09/06/08/BuildingRigorousEvidencetoDrivepolicy>> (consulté le 2 mai 2014).
- PECH M.-E., « Nicolas Sarkozy inaugure un «internat d’excellence» », *Le Figaro*, 9 septembre 2010.
- SIMON C., « Esther Duflo: Des neurones contre la misère », *Le Monde*, 2009.
- VALDENNAIRE M., BERARD J., « L’expérimentation pour renouveler les politiques publiques ? », La Vie des idées (25 juin 2013), en ligne : <<http://www.laviedesidees.fr/L-experimentation-pour-renouveler.html>> (consulté le 18 septembre 2013).

Actes de colloques

- CHOPART J.-N., MARTIN C., « Sciences de l’Homme et de la société et évaluation des politiques publiques », in *Evaluation et Décision, séminaire de méthode*, coll. Plan Urbain, CEOPS, Ecole Nationale des Travaux Publics de l’Etat, 1990, p. 139- 149.
- CLEMENS M., « Applying evaluation to development and development aid », in *Malaise dans l’évaluation: quelles leçons tirer de l’expérience du développement?*, 9ème conférence AFD/EUDN, Paris, 26 mars 2012.
- CNFPT, « Les dossiers documentaires des Conférences, Vers un revenu de solidarité active ? » Conférence du 5 décembre 2007, CNFPT première couronne, 5 décembre 2007.
- CRASNOW S. « Evidence for Use: The Role of Case Studies in Political Science Research », Minnesota, SPSP 2009: Society for Philosophy of Science in Practice, 18 juin 2009.

- DARES, « Colloque sur les expérimentations pour les politiques publiques de l'emploi et de la formation », Paris, 2008, p. 168.
- DIRECTION GENERALE DE LA COHESION SOCIALE, ANSA. « L'expérimentation sociale: innover par la preuve? », Centre d'Analyse Stratégique, 25 novembre 2010.
- DUFLO E., « Introduction », in *Conférence nationale de l'expérimentation sociale, réformes contre placebo?*, Paris, 2010.
- GURGAND M., « Les méthodes d'évaluation statistique des politiques publiques », IRESO, 2004.
- MUCCHIELLI A., « Les processus intellectuels fondamentaux sous-jacents aux techniques et méthodes qualitatives », in *Bilan et perspectives de la recherche qualitative en sciences humaines et sociales*, 3, Université Paul Valéry, Montpellier III, Collection hors série « Les actes », 2006.
- PICAVET O., « Expérimenter ou évaluer les politiques publiques », séminaire praticiens/Chercheur, ICPDE/SGMAP, 16 novembre 2013
- ROCARD M., « Les nouvelles frontières de l'évaluation », Marseille, 2009.
- RODRIK D., « The new development economics: We shall experiment, but how shall we learn? », *Brookings Development Conference*, 29 mai 2008, p. 14.

Table des figures et graphiques

Figure 1: La double logique de l' <i>Evidence-Based Policy</i>	58
Figure 2: Logique de l'analyse d'attribution.....	75
Figure 3: Logique de l'analyse de contribution	76
Figure 4: Les quatre genres identifiés dans la littérature sur l'apprentissage politique. 110	
Figure 5: Les quatre espèces d'apprentissages du genre « apprentissage épistémique »	111
Figure 6: Logique de l'induction analytique.....	121
Figure 7 : Représentation réticulaire du champ d'action stratégique émergent de l'expérimentation par assignation aléatoire de 2005 à 2007 (voir document agrandi en annexe 3).....	160
Figure 8: Représentation réticulaire avec calcul de « <i>average shortest path length</i> » du champ d'action stratégique émergent de l'expérimentation par assignation aléatoire de 2005 à 2007 (voir document agrandi en annexe 3),.....	161
Figure 9 : Représentation réticulaire du champ d'action stratégique émergent de l'expérimentation par assignation aléatoire de 2007 à janvier 2009 (voir document agrandi en annexe 3).....	188
Figure 10: Représentation réticulaire du champ d'action stratégique émergent de l'expérimentation par assignation aléatoire de 2007 à janvier 2009, présentant les individus ayant au moins 4 connections avec le reste des individus du réseau (voir document agrandi en annexe 3).....	189
Figure 11: Représentation réticulaire avec calcul de « <i>average shortest path length</i> » du champ d'action stratégique émergent de l'expérimentation par assignation aléatoire de 2007 à janvier 2009 (voir document agrandi en annexe 3),.....	190
Figure 12: Procédure de déroulement des appels à Projets au 1er semestre 2010	195
Figure 13: Représentation des liens entre évaluateurs par expérimentation (rouge) et par appartenance de laboratoire (bleu) de l'École d'Économie de Paris et du Crest	243
Figure 14: Représentation graphique de l'ingénierie de l'expérimentation "Mallette des parents" et du nombre de bénéficiaires	274
Figure 15: Ingénierie du tirage aléatoire pour l'expérimentation "inscrire les contrats en alternance dans une logique de parcours sécurisé », porté par la Mission Locale de Beaune	275

Figure 16: Les différents types d'interventions et nombre de bénéficiaires des neuf groupes de l'expérimentation Rebond.....	281
Figure 17: Les champs d'action stratégiques (CAS) influençant le CAS de l'expérimentation sociale par assignation aléatoire.....	318

Graphique 1 : Nombre d'expérimentations sociales évaluées par assignation aléatoires initiées entre 1961 et 1995 (source : Greenderg et Shroder 1997).....	47
Graphique 2 : Nombre d'expérimentation en laboratoire (laboratory experiments) et d'expérimentations de terrain (field experiments) publiées dans les cinq principales revues académiques en économie de 1975 à 2010.....	53
Graphique 3 : Evolution du nombre d'évaluations d'impact à la Banque Mondiale utilisant la méthode expérimentale par assignation aléatoire (experimental) et quasi-expérimentale initiées entre 2000 à 2010 (source)	70
Graphique 4: Évolution de la provenance des Fonds de PSE, en Kilo euros (2008-2011)	156
Graphique 5:Nombre de projets évalués par la méthode d'évaluation aléatoire parmi les appels à projets dont l'évaluation n'était pas centralisée au niveau national (période 2009-2013).....	206
Graphique 6: Montant des fonds engagés pour les évaluations et pour les projets par le FEJ en million d'euros pendant ses différentes périodes.....	217
Graphique 7: Nombre d'évaluation aléatoires élaborées par an sur le territoire français	222

Table des tableaux et encadrés

Table 1 : Types de pratiques de travail institutionnel identifiés par T. Lawrence et R. Suddaby	92
Table 2 : Typologie d'acteurs selon leurs compétences sociales et la nature de leur travail institutionnel.....	98
Table 3 : Tableau récapitulatif de l'ensemble des cas étudiés dans ce chapitre	231
Table 4: Présentation de la diminution des échantillons pour les15 expérimentations étudiées.....	293

Encadré 1 : "The lady tasting tea" la logique expérimentale proposée par R. A. Fisher.....	40

Encadré 2 : Critique de D. Campbell sur l'utilisation des expérimentations aléatoires pour l'évaluation des politiques publiques.....	49
Encadré 3 : Protocole des revues systématiques proposé par la Collaboration Cochrane	59
Encadré 4: Documents produits par le ministère de l'éducation américain, relatifs aux méthodes d'évaluation jugées comme "scientifiques".....	65
Encadré 5 : Communiqué de l'association américaine d'évaluation « Scientifically Based Evaluation Methods ».....	72
Encadré 6: Martin Hirsch, un entrepreneur aux multiples identités.....	146
Encadré 7: La Délégation Interministérielle à l'Innovation, à l'Expérimentation Sociale et à l'Economie Sociale.	152
Encadré 8: La genèse de l'école de l'économie de Paris, Paris School of Economics...	155
Encadré 9: Avis du Comité économique et social européen sur « Comment l'expérimentation sociale peut elle servir en Europe à l'élaboration des politiques publiques.....	183
Encadré 10: Composition du Conseil de Gestion interministériel	196
Encadré 11: Composition du Conseil Scientifique du Fonds d'expérimentation pour la jeunesse.....	198

Table des annexes

ANNEXE 1 : PRESENTATION DU CABINET D’EVALUATION EUREVAL, CO-FINANCEUR DU CONTRAT CIFRE	352
ANNEXE 2 : LISTE DES ENTRETIENS REALISES	353
ANNEXE 3 : ANALYSES DE RESEAU.....	357
ANNEXE 4 :MEMBRES DU CABINET DU HCSA ET HCJ	362
ANNEXE 5 : TAILLE D’ECHANTILLON ET PUISSANCE STATISTIQUE	363

Annexe 1 : Présentation du cabinet d'évaluation Euréval, co-financeur du contrat Cifre



Evaluation des politiques publiques : un savoir-faire au service de la décision et de l'action publique

Euréval a été créé à Lyon, dès 1994, sous le nom de Centre Européen d'Expertise en Évaluation (CEE). Ses fondateurs, Eric MONNIER et Jacques TOULEMONDE sont deux experts issus de la recherche dans le domaine de l'évaluation des politiques et programmes publics en Europe et en France. En 2001, une équipe, basée à Paris, a permis d'accompagner le développement des commandes publiques.

Reconnue comme l'une des toutes premières équipes européennes dans son domaine, **EURÉVAL** est forte, aujourd'hui, d'une **quinzaine d'intervenants** issus de disciplines diverses : sociologie, sciences politiques, économie, gestion et sciences de l'ingénieur.

Euréval accompagne et conseille les administrations et services publics à tous les niveaux, depuis les collectivités territoriales (communes, structures intercommunales, Conseils généraux et régionaux) et administrations nationales (ministères, agences), jusqu'aux institutions internationales (Commission européenne, Banque mondiale, etc.).

Ses prestations incluent :

- L'évaluation de politiques, programmes ou actions publics ainsi que l'assistance aux organisations qui commandent, pilotent et utilisent l'évaluation ;
- Le conseil stratégique et le développement de systèmes de pilotage de la performance ;
- La formation de gestionnaires et de décideurs publics, notamment par des formations actions en interne et dans le cadre de séminaires sur catalogue, à Paris et Bruxelles.

Dotée d'une expertise sectorielle, Euréval est structurée autour de trois pôles, Politiques Économiques dirigé par Virginie Besrest, Santé & Solidarités dirigé par Hélène Genin et Environnement.

NOS ENGAGEMENTS

- ♦ **Écoute et médiation**
- ♦ **Pluridisciplinarité**
- ♦ **Qualité & professionnalisme**
- ♦ **Indépendance**

« Au service de l'action publique »

NOS RESEAUX

Euréval est membre fondateur de :





Société Française de l'Évaluation

11 bis Place Julien Ferry
69005 LYON
10 rue du Château d'eau
75010 PARIS
Tél : 04 72 83 75 00
lyon@eureval.fr
www.eureval.fr

3 pôles de compétences

Pôle Politiques Economiques :

Champs d'intervention :

- Aides au développement des entreprises
- Recherche, innovation et transfert de technologie
- Développement économique régional, attractivité
- Responsabilité sociale des Entreprises

Exemples de références :

- Actions collectives : Evaluation du dispositif régional de soutien à la compétitivité des entreprises Rhône-alpines et Evaluation des actions collectives du Ministère de l'Industrie
- Evaluation de la politique de soutien à la création-transmission d'entreprises pour le Conseil régional du Limousin

Politiques Santé & Solidarités

Champs d'intervention :

- Les politiques sociales (personnes âgées, enfance/famille, handicap)
- Les politiques régionales de santé publique
- Les actions de prévention de santé
- Lutte contre les discriminations et égalité des chances

Exemples de références :

- Evaluation des Plans Régionaux de Santé Publique (Bourgogne, Rhône Alpes, Centre, Bretagne)
- Evaluation de la politique en faveur des personnes âgées menée par le Conseil Général du Loiret

Pôle Politiques Environnementales

Champs d'intervention :

- Evaluations Stratégiques Environnementales
- Politiques contractuelles dans le domaine de l'énergie et de l'environnement
- Formation
- Politiques éco-environnementales

Exemples de références :

- Evaluation Stratégique Environnementale du CPER 2007-2013 pour le SGAR Picardie
- Evaluation du soutien à l'export des éco-entreprises

De nombreuses missions européennes

Domaines :

- Politiques de coopération et aide au développement
- Politique Agricole Commune
- Programme pour les médias et l'audiovisuel
- Politiques de formation professionnelle

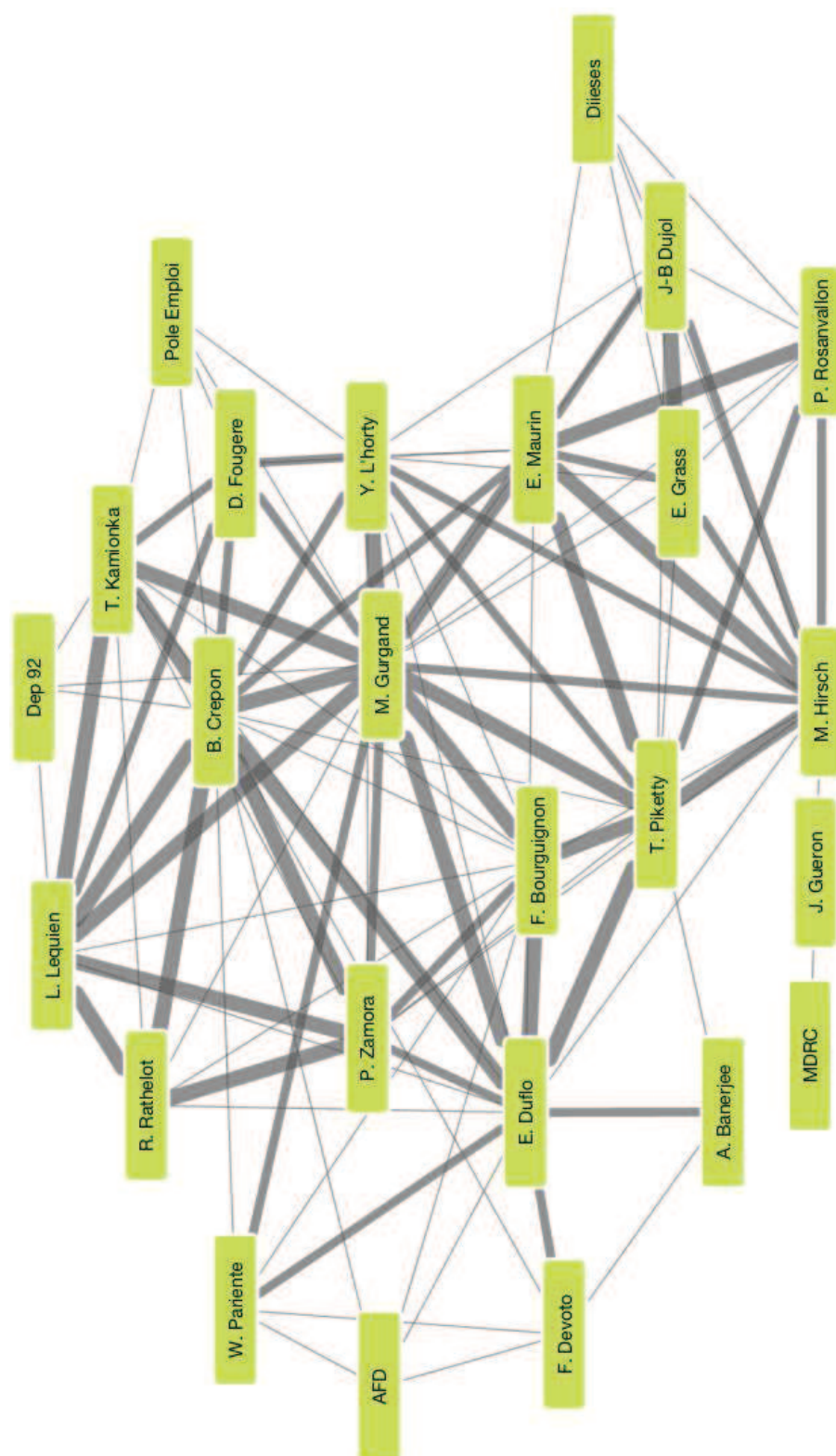
Annexe 2 : Liste des entretiens réalisés

Institution	Poste	Mode d'entretien	Durée de l'entretien
Fonds d'expérimentation pour la Jeunesse	Responsable du pôle évaluation/docteur en économie	Face-à-face	1h30
FEJ	Responsable de la capitalisation/docteur en histoire	Face-à-face	1h
FEJ	Ex Chef de la Majef	Face-à-face	1h30
Pôle Emploi	Direction Veille, prospective et Affaires internationales	Face-à-face	1h30
	Sous directeur Animation des Evaluations, Veille et Benchmark		
Haut commissariat aux solidarités actives et à la jeunesse	Haut Commissaire	Face-à-face	1h
Cabinet des Hauts Commissariats (HC)	Conseiller technique du fond d'expérimentation/Administrateur de l'insee/Doctorant PSE	Face-à-face	1h30
Cabinet (HC)/Mafej	Stagiaire/doctorant PSE	Face-à-face	1h30
Cabinet (HC)	Directeur adjoint	Face-à-face	1h
Cabinet (HC)	Conseillère technique évaluation	Téléphone	1h30
Ansa	Chargé de mission	Face-à-face	1h
SGMAP	Service évaluation des politiques publiques	Face-à-face	1h
Direction générale de la modernisation de l'Etat	Chef de projet innovation	Face-à-face	30 min
Ville de Grenoble	Adjoint à l'action sociale/Missionné par M. Hirsch d'une étude sur les expérimentations	Face-à-face	30 min
CFA les 13 vents	Directeur	Téléphone	1h
Mission Locale de	Responsable Secteur Développement, Partenariat, Relais de Proximité à la	Téléphone	1h30

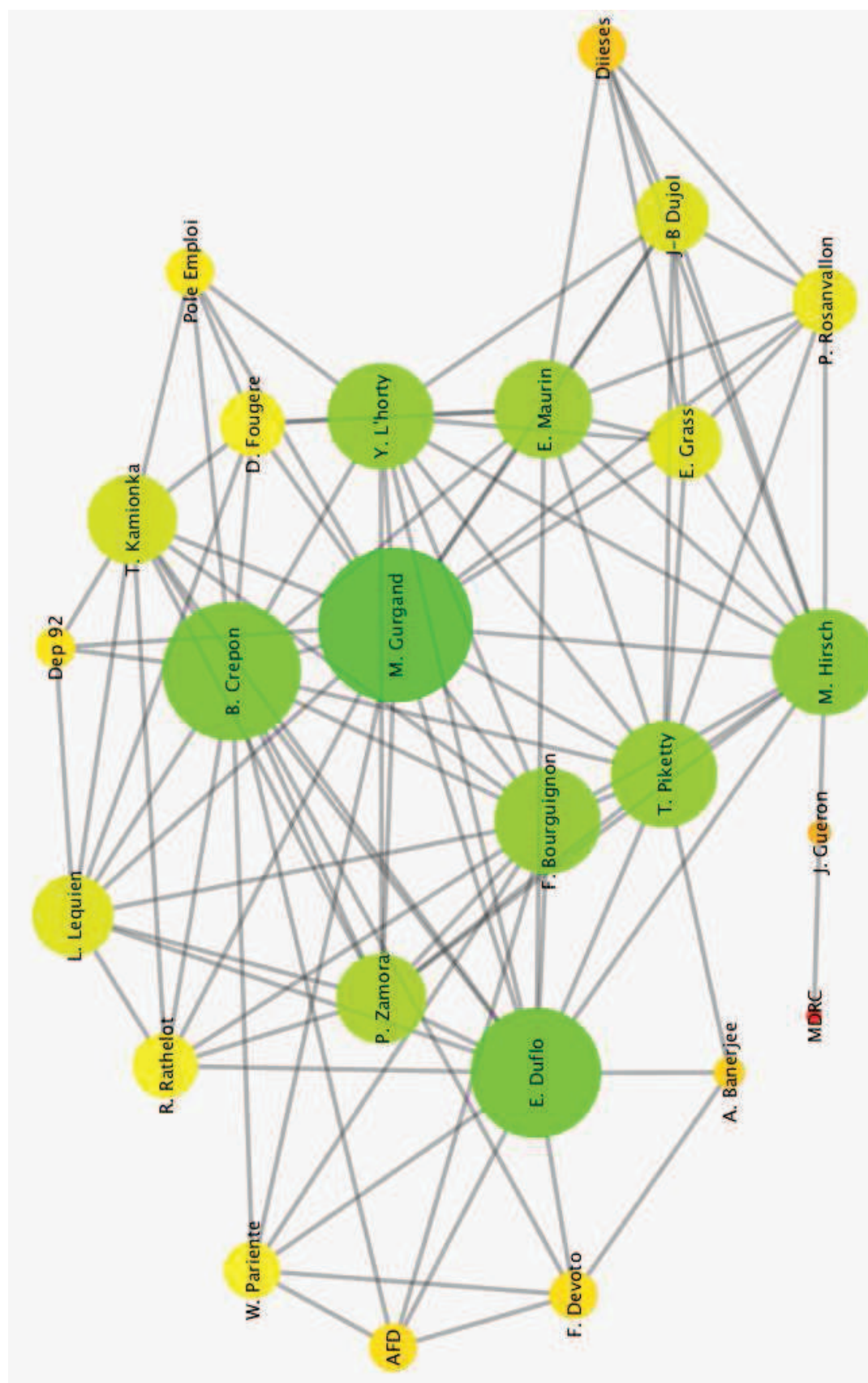
Marseille	Mission Locale de Marseille				
Mission Locale de Tulle	Chargée de projet	Dispositif	Local	Téléphone	1h30
Mission Locale de Tulle	Ex directeur			Face-à-face	2h
AFD	Chargée de l'évaluation			Face-à-face	2h
Euréal	Directeur			Face-à-face	1h30
Ministère de l'éducation	Ex Directeur général de l'enseignement scolaire			Face-à-face	1h
Ministère de l'éducation	chef du département recherche-développement, innovation et expérimentation à la DGESCO			Face-à-face	1h
Crest, Dress	Administrateur, chef du bureau des retraites			Face-à-face	1h
Dares	Ex sous-directrice du suivi et de l'évaluation des politiques de l'emploi et de la formation professionnelle			Face-à-face	1h30
Dares	Ex directrice			Face-à-face	1h30
Dares	Ex directeur			Face-à-face	1h
Dares	Directrice			Face-à-face	1h
Dares, Crest, Insee	Administrateur hors classe			Face-à-face	1h30
Crest, Dares	Administrateur			Face-à-face	30 min
Crest, Dares, J-pal Europe	Administrateur, J-pal Europe			Face-à-face	1h30
J-Pal Europe	Directeur-adjoint du Bureau Europe			Face-à-face	2h
Fédération TEPP	Maître de conférences			Face-à-face	1h30
Fédération TEPP	Professeur, directeur de fédération de recherche			Face-à-face	1h30
CREST/J-pal Europe	Administrateur hors classe			Face-à-face	1h30
Crest	Directeur de recherche CNRS				1h
Centre pour (CEE)	d'Etude l'Emploi	Chargé de recherche CNRS		Face-à-face	3h30

PSE/INRA/CREST/ J-pal Europe	Chargé de recherche	Face-à-face	1h
J-pal Europe PSE, LIEPP	Chercheur associé	Téléphone	1h
J-pal Europe PSE	Directeur de recherche CNRS	Téléphone	1h
J-pal Europe PSE	Post doc	Face-à-face	1h
Centre d'économie de la Sorbonne	Directeur de recherche CNRS	Face-à-face	1h30
Membre fondateur SFE	Professeur	Face-à-face	1h30
Centre associé Céreq/Iredu	Directeur de recherche émérite au CNRS	Téléphone	1h30
Céreq	Chargée d'études	Téléphone	1h30
Céreq	Chargée d'études	Téléphone	1h30
Céreq	chargée de mission pour les Partenariats nationaux	Téléphone	1h30
Credoc	Directrice adjointe du département évaluation des politiques sociales	Téléphone	1h30

Annexe 3 : Analyses de réseau

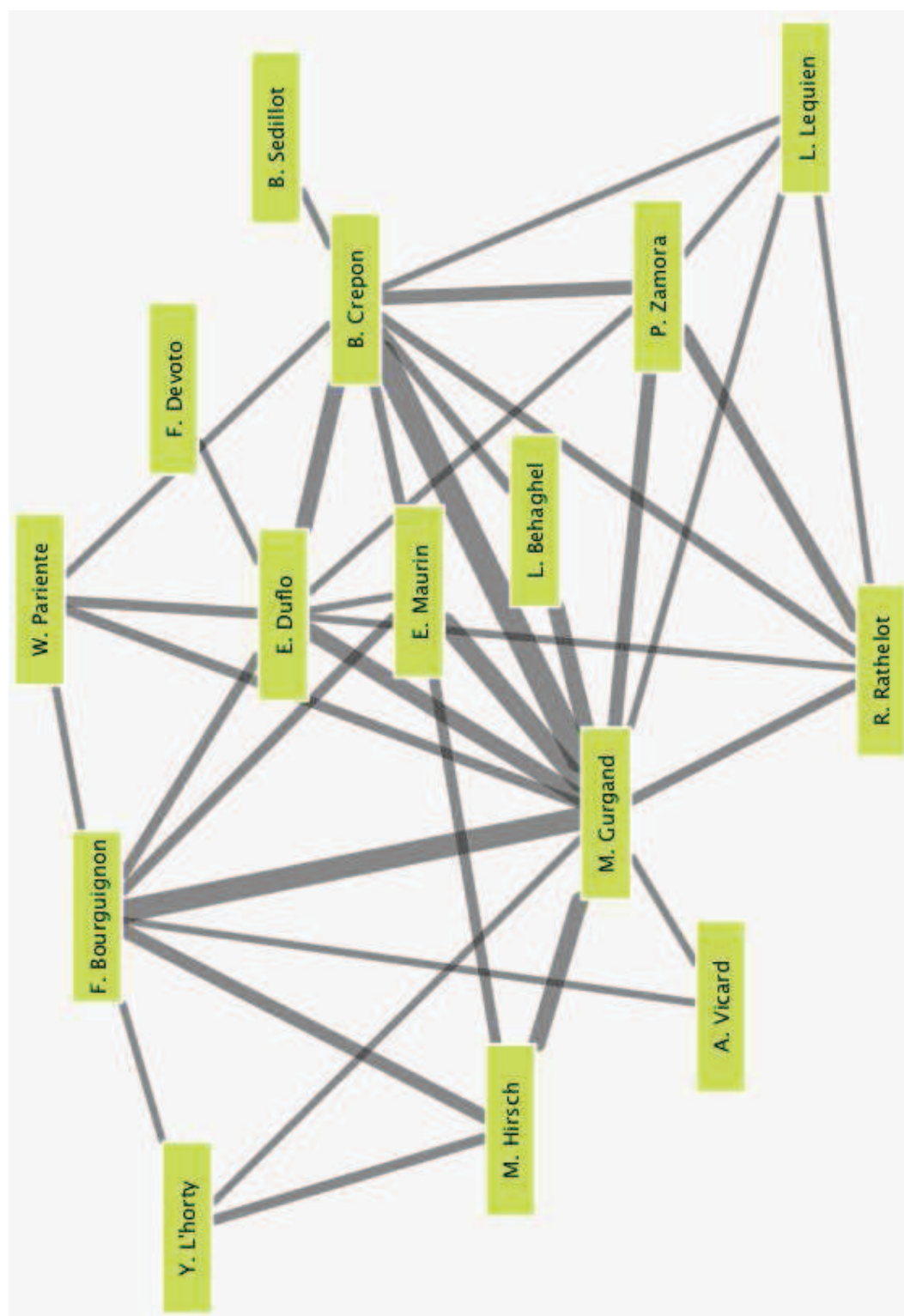


Représentation réticulaire du champ d'action stratégique émergent de l'expérimentation par assignation aléatoire de 2005 à 2007



Représentation réticulaire avec calcul de « average shortest path length » du champ d'action stratégique émergent de l'expérimentation par assignation aléatoire de 2005 à 2007

DEVAUX-SPATARAKIS Agathe | Thèse pour le doctorat en Science politique | 2014



Représentation réticulaire du champ d'action stratégique émergent de l'expérimentation par assignation aléatoire de 2007 à janvier 2009, présentant les individus ayant au moins 4 connections avec le reste des individus du réseau

DEVAUX-SPATARAKIS Agathe | Thèse pour le doctorat en Science politique | 2014

Annexe 4 : Membres du Cabinet du HCSA et HCJ¹⁰⁰²

Nom	Position
Emmanuelle WARGON	Directrice de Cabinet, Nommée le 22 mai 2007 ; Même poste sous le HCJ
Marianne DURANTON	Chef de cabinet (2007- avril 2008), Conseillère technique, Relations avec le Parlement (avril 2008) ; Même poste sous le HCJ
Étienne GRASS	Chargé de mission (2007-2008), Conseiller technique en charge du revenu de solidarité active (avril 2008) ; Directeur adjoint du cabinet (juillet 2008). ; Directeur de cabinet adjoint à la jeunesse (juillet 2009)
Jean-Benoît DUJOL	Chargé de mission (2007- 2008) ; Conseiller technique, Revenu de solidarité active (avril 2008) ; Directeur adjoint du cabinet (juillet 2009).
Virginie MAGNANT	Conseillère technique, Questions budgétaires (juillet 2007-août 2009) Même poste sous le HCJ
Marie-Odile SIMON	Conseillère technique, Évaluation (juillet 2007-juin 2009)
Laure KERMEN-LECUIR	Conseillère technique, Suivi des expérimentations/départements Même poste sous le HCJ
Jean-Philippe DANIEL	Chef de cabinet (avril 2008) ; Même poste sous le HCJ
Patrick CHANSON	Conseiller technique, Communication depuis avril 2008 ; Même poste sous le HCJ
Nathalie SIPRES	Conseillère technique, Questions d'emploi (juillet 2008) (n'apparaît pas)
Séverine FRANCOIS	Adjointe du chef de cabinet
Yann DEBOS	Conseiller technique chargé des politiques de l'emploi (départ sept 2009)
Loïc DUROSELLE	Conseiller technique chargé de la jeunesse et du service civique départ aout 2009
Augustin VICARD	Conseiller technique chargé du fonds d'expérimentations (juillet 2009) en remplacement de Marie-Odile Simon
Jérôme TEILLARD	Conseiller technique chargé de l'orientation ; Conseiller technique orientation et insertion professionnelle (oct 2009)
Catherine PALPANT	Conseillère technique chargée des questions européennes et internationale
Sophie BENTEGEAT	Conseillère technique chargée de la vie associative
Loïc LE NAOUR	Chargé des interventions
Elise HERMANT	Chargé de mission communication
David GRUSON	Chargé de mission jeunesse
Yannick MIEL	Chargé de mission jeunesse
Inès MIMIN	Conseillère technique pôle jeunesse octobre 2009
Lionel LEYCURAS	Conseiller technique pôle jeunesse (remplace Loïc Duroselle) octobre 2009
Xavier MOAL	Conseiller budgétaire remplace virginie Magnant (oct 2009)
Malissa MARSEILLE	conseillère technique RSA (oct 2009)

¹⁰⁰² ACTEURS PUBLICS, « Le cabinet de Martin Hirsch », *acteurspublics.com* (23 juillet 2008), en ligne : <<http://www.acteurspublics.com/2008/07/23/le-cabinet-de-martin-hirsch>> (consulté le 29 juillet 2013). Martin HIRSCH, *Arrêté du 10 avril 2008 portant nomination au cabinet du haut-commissaire aux solidarités actives contre la pauvreté*, 10 avril 2008. Martin HIRSCH, *Arrêté du 1er juillet 2008 portant nomination au cabinet du haut-commissaire aux solidarités actives contre la pauvreté*, (2008), JORF n°0160.

Annexe 5 : Taille d'échantillon et puissance statistique

Encadré 3. Taille d'échantillon et puissance statistique

Au moment d'élaborer une méthodologie d'enquête quantitative, une question revient en permanence, celle de la **taille minimale que doit avoir l'échantillon** de personnes interrogées pour obtenir des résultats « suffisamment précis ». Ce « suffisamment précis » dépend non seulement de la taille de l'échantillon mais aussi la valeur atteinte par l'indicateur mesuré.

Dans le cas d'une expérimentation randomisée, la situation de départ des membres du groupe bénéficiaire (B) est en moyenne la même que celle du groupe témoin (T) puisque les deux groupes ont été constitués de façon aléatoire. Il s'agit de voir si en fin d'expérimentation leur situation est significativement différente ou non.

L'intervalle de confiance (IC) d'un écart entre deux proportions p_1 et p_2 , estimées sur des populations de taille respective n_1 et n_2 , est donné par la formule suivante :

$$IC = \hat{p}_1 - \hat{p}_2 \pm 1,96 \sqrt{\frac{\hat{p}_1(100-\hat{p}_1)}{n_1} + \frac{\hat{p}_2(100-\hat{p}_2)}{n_2}}$$

Cette formule fonctionne dans le cas d'une sélection venant d'un sondage aléatoire simple, ce qui n'est pas toujours le cas. Certaines modalités d'enquête peuvent améliorer la situation présente (tirage stratifié, sondage équilibré, calage) ou la détériorer (tirage à plusieurs degrés, par exemple si l'on tire des zones et des individus dans la zone, on a alors de l'effet de grappe). Cette formule permet néanmoins de fournir un scénario central et donne des ordres de grandeur relativement robustes.

La formule correspond à un intervalle de confiance avec une probabilité de 95%, c'est-à-dire que dans 5 % des cas la différence des proportions réelles sortira de cet intervalle de confiance. Autrement dit, si l'on décide que les deux populations sont différentes à partir du moment où l'IC ne comprend pas la valeur 0, on aura tort dans moins de 5% des cas (cas où 0 est une des bornes de l'IC).

Le tableau 2 donne la valeur de n ($n_1=n_2$) telle que $IC > 0$ (dans le cas où $\hat{p}_1 > \hat{p}_2$).

Tableau 2 : Taille d'échantillon théoriquement requise pour chaque sous-population B et T, selon le degré de précision souhaité et l'ampleur du phénomène analysé

Valeur de l'indicateur mesuré	Degré de précision souhaité					
	1 point	2 points	3 points	4 points	5 points	10 points
5% (ou 95%)	4 000	1 100	550	350	250	100
10% (ou 90%)	> 8 000	2 000	900	550	350	100
15% (ou 85%)	> 8 000	2 700	1 200	700	450	150
20% (ou 80%)	> 8 000	3 200	1 500	850	550	150
25% (ou 75%)	> 8 000	3 700	1 700	950	650	200
30% (ou 70%)	> 8 000	> 8 000	1 850	1 050	700	200
35% (ou 65%)	> 8 000	> 8 000	2 000	1 150	750	200
40% (ou 60%)	> 8 000	> 8 000	2 150	1 200	750	200
45% (ou 55%)	> 8 000	> 8 000	2 150	1 200	800	200
50%	> 8 000	> 8 000	2 150	1 200	800	200

Lecture : on peut considérer qu'un écart de 5 points (où l'indicateur analysé vaut 20% (ou 80%) dans l'un des groupes et 25% (ou 85%) dans l'autre) est statistiquement significatif (avec une probabilité de 95%) à partir d'un échantillon total minimum de 550 bénéficiaires et 550 témoins répondants. En dessous de cette taille d'échantillon, cet écart n'est pas significatif. Pour qu'un écart de 2 points soit significatif quand l'indicateur analysé prend des valeurs plus élevées, la taille de l'échantillon doit être augmentée.

source : Isa ALDEGHI, Pauline JAUNEAU et Lara MULLER, *Bilan méthodologique des évaluations conduites par le Crédoc dans le cadre du Fonds d'expérimentation pour la jeunesse*, 300, coll. Cahier de recherche, CREDOC, 2012, p. 36.