



DOCTORAT AIX-MARSEILLE UNIVERSITE

Délivré par Aix-Marseille Université

Creative product assessment in design:

Influence of judges' backgrounds and levels of experience in design

THESE

présentée pour l'obtention du grade de

DOCTEUR D'AIX-MARSEILLE UNIVERSITE

Champ disciplinaire : Psychologie Cognitive et Ergonomique

par

Alicja WOJTCZUK

Composition du jury

Examinatrice : Marie-Laure BARBIER, HDR, Maître de Conférences, Centre de Recherche en Psychologie de la Connaissance, du Langage et de l'Emotion, Aix-Marseille Université

Directrice : Nathalie BONNARDEL, Professeur des Universités, Centre de Recherche en Psychologie de la Connaissance, du Langage et de l'Emotion, Aix-Marseille Université

Rapporteur : Carole BOUCHARD, Professeur, Design and Innovation Laboratory (LCPI), Arts et Métiers Paristech

Examineur : Karl J. JEFFRIES, Senior Lecturer, School of Art, Design and Performance, University of Central Lancashire, UK

Rapporteur : Todd LUBART, Professeur, Université Paris Descartes - Paris V

Formation Doctorale Cognition, Langage, Education – ED356

ACKNOWLEDGMENTS

I express my gratitude and sincere consideration for all persons who have helped and supported me during the completion of this work. My special thanks go to:

Nathalie Bonnardel, my research director, who supervised my work since my Master degree. I want to thank you for giving me the chance to count amongst your PhD students, for all the help, time and understanding I could receive from your side.

Members of the jury for agreeing to evaluate this PhD thesis. I am honored. My first thanks go to Mrs Carole Bouchard and Mr. Todd Lubart, the reporters of this dissertation. I also thank Mrs Marie-Laure Barbier and Mr Karl Jeffries for agreeing to participate in the jury of my Phd.

Moreover, I would like to thank all the researchers and colleagues that I encountered when working on different research projects, especially to Julien Nelson, Marion Botella, Thomas Arciszewski, Marcin Smietana, Haydee Carrasco-Ortiz and Vlad Petre Glavenau, for their precious advices.

All the design professionals and students, who participated to my studies and shared knowledge about their professional lives and opinions. Special thanks to René Ragueb et Marylaine Pupet, for your time and energy. With your help the collaboration with design schools and professionals was much more easy.

Thanks to Carlos Azevedo, for time necessary to start this thesis and to Arnoud Boekhoorn, for time necessary to finish it.

Laurent Olivares, thank you for all the psychological and practical support.

Finally, thanks to family, friends and colleagues for their encouragements during this long journey.

RÉSUMÉ

L'objectif général de cette thèse est d'étudier l'évaluation de la créativité dans le domaine du design. Plus précisément, ce travail se centre sur l'évaluation de productions créatives en design graphique et il repose sur des méthodes complémentaires d'analyse des critères utilisés par des juges. L'approche adoptée vise à explorer les variations de jugements, dans différents contextes d'évaluation, afin d'identifier les facteurs influant sur les critères (ou les « référents évaluatifs ») qui sont pris en considération par les juges. Deux facteurs ont plus particulièrement été pris en compte : les points de vue adoptés par les juges en fonction de leur parcours professionnel (designers, directeurs artistiques, enseignants en design et public visé) et le niveau d'expertise en design (experts affirmés, experts intermédiaires et non-experts).

Cette recherche a été réalisée dans différents contextes d'évaluation et elle tente d'identifier les éléments caractéristiques des jugements, selon le profil des juges émettant ces jugements. La première étude vise à comprendre les représentations mentales des juges, en explorant les critères qu'ils déclarent importants pour la créativité. La seconde étude analyse les corrélations entre les scores attribués aux productions en design sur un ensemble des critères, ainsi que les niveaux d'accords inter-juges pour chacun de ces critères. La troisième étude permet une analyse qualitative des verbalisations spontanées exprimées par des juges durant leurs évaluations de productions en design.

MOTS CLÉS

Créativité ; Design graphique ; Evaluation de créativité ; Critères d'évaluation ;
Profils des juges

ABSTRACT

The present thesis aims to study creativity assessments in design. More precisely, this research focuses on the evaluation of creative productions in graphic design area and it is based on complementary methods of analysis of used criteria and “evaluative referents”. It aims to identify, in various assessment contexts, factors that exert an influence on the judgments of creative productions. It develops a multiple feedback approach by exploring assessments made by judges with different professional backgrounds (designers, art directors, design teachers and targeted audience) and levels of experience in design (asserted experts, intermediary experts and laypeople).

The research frame includes different contexts of assessment situations and tries to capture the characteristics of judges’ approaches to creativity in design, on the basis of three complementary studies. The first study focusses on judges’ mental representations by exploring criteria they declare important to creativity in design. The second study allows an analysis of correlations between scores attributed to design productions with regard to different criteria as well as an analysis of inter-judge agreement on them. The third study aims to perform qualitative analyses of spontaneous verbalizations expressed by judges during their analyses of design outcomes.

KEYWORDS

Creativity; Graphic design; Creativity assessment; Assessment criteria; Judges profiles

RÉSUMÉ SUBSTANTIEL

L'étude de la créativité présente un intérêt majeur dans le contexte actuel de notre société (Andiliou & Murphy, 2010). Il est primordial pour les entreprises de déterminer dans quelle mesure leurs produits peuvent être considérés comme créatifs, car l'innovation peut leur permettre de générer 75% de leurs revenus (Milton, 2003). De ce fait les entreprises mais aussi les professionnels chargés d'exprimer leur potentiel créatif, ont besoin de recourir à des méthodes et/ou à des mesures permettant de distinguer les produits qui sont considérés comme les plus créatifs de ceux qui le sont moins.

L'objectif général de cette thèse est de contribuer à l'analyse des différents facteurs responsables de différences dans la manière dont les travaux créatifs sont évalués. Nous avons choisi de focaliser notre recherche sur le domaine du design graphique, décrit par La Maison de Artistes comme destiné à « transmettre un message visuel dans tous les domaines de la vie ». Réduit aux deux dimensions, il a principalement son rôle d'informer et de construire l'image de marque. Son côté fonctionnel est réduit, comparé au design produit, lui concernant des objets tridimensionnels, centrés sur les usages et la manipulation tactile. Notre but est d'approfondir les connaissances quant aux points communs et aux différences entre le domaine du design graphique et celui du design produit. Les deux domaines étant omniprésents dans la vie quotidienne, ce dernier a pu bénéficier d'être l'objet de bien plus d'études portant sur la créativité.

Plus précisément, notre objectif général peut être subdivisé en deux parties :

- Premièrement, nous voulons identifier les critères utilisés spontanément par les juges durant l'évaluation du caractère créatif de productions relevant du design graphique, et les comparer à ceux utilisés pour évaluer la créativité de productions relevant du design produit. De plus, nous voulons déterminer dans quelle mesure les critères déclarés comme importants par les juges - sur la base de leurs représentations mentales concernant la créativité en design - sont différents des critères appliqués lors de l'attribution de scores de créativité et de ceux spontanément verbalisés dans des situations d'évaluation.
- Deuxièmement, nous souhaitons déterminer les différences entre les évaluations réalisées par des juges ayant des profils différents. Les profils pris en compte reposent sur deux facteurs : le « point de vue professionnel » des juges - dépendant de leur statut

professionnel et pouvant influencer leur perception de produits design - et leur niveau d'expérience dans le domaine du design (mesuré en nombre d'années passées à exercer une activité dans ce domaine).

Le contexte induit par le but ou par la nature de l'évaluation peut exercer une influence sur les résultats qui seront obtenus. Aussi, afin de contribuer à une meilleure compréhension de la façon dont la créativité des productions design peut être évaluée, nous avons mis en œuvre différentes méthodes d'analyse :

- 1) Tout d'abord, nous avons pris en compte les représentations mentales que les différents juges peuvent avoir quant à la créativité dans le domaine du design graphique.
- 2) Nous avons ensuite analysé la manière d'attribuer des scores à différentes productions dans ce domaine, en imposant aux participants la prise en compte de critères de créativité bien définis.

Finalement, afin d'identifier plus précisément, les sources de variation intervenant dans les évaluations de la créativité, nous avons analysé les verbalisations spontanées produites par les juges durant une situation réelle d'évaluation des produits relevant du design graphique. Dans ce dernier cas, nous avons pris en compte les évaluations basées sur leur première impression, celles reposent sur une analyse approfondie de la créativité des œuvres, et celles exprimées en fonction de préférences personnelles.

1.1 Partie 1 - Etat de l'art

1.1.1 Créativité

La créativité, avec toute sa complexité et ses aspects multidimensionnels, est difficile à définir de manière claire et facilement généralisable. Pour ce travail de recherche, nous allons nous référer à la définition proposée par Sarkar et Chakrabarti (2008), issue de leur analyse d'environ 160 définitions issues de diverses recherches dans le domaine de créativité. Selon eux, le point commun à toutes les définitions de la créativité est qu'elle apparaît à travers un processus mis en œuvre par l'individu pour générer des idées, des solutions ou des produits, qui sont à la fois nouveaux et valables (adaptés aux contraintes).

Si l'on considère le modèle des quatre P (Rhodes, 1961; Runco, 2004), les études portant sur les activités de conception créative (dont le design) peuvent prendre en compte quatre paramètres : le concepteur (« Person »), le processus de conception (« Process »), les productions (« Products ») et l'environnement matériel et social (« Place »). Notre recherche se focalise clairement sur le caractère créatif des productions design, celles-ci pouvant être considérées comme des objets à la fois innovants et adaptés aux contraintes et au contexte auquel ils sont destinés (Bonnardel, 2006; Bonnardel, 2009; Bonnardel & Zenasni, 2010; Sternberg & Lubart, 1999). Néanmoins, il est important de savoir que les quatre paramètres évoqués ci-dessus interagissent entre eux. Ainsi, une production résulte d'un processus créatif, qui est lui-même associé, en psychologie cognitive, à une activité de résolution de problème à la fois « ouvert » - admettant un grand nombre des solutions possibles (Fustier, 1989) - et « mal structuré », la représentation initiale du concepteur étant imprécise et incomplète (Eastman, 1969; Reitman, 1964, Simon, 1973, 1995). De plus, ce résultat est produit par une personne présentant un certain nombre de caractéristiques cognitives et conatives qui interagissent dans un environnement physique et social. Ces différents facteurs ont alors un impact plus ou moins favorable à la créativité (Lubart, Mouchiroud, Tordjman & Zenasni, 2003 ; Sternberg & Lubart, 1995).

En outre, afin d'être reconnues comme créatives, les productions doivent être jugées par ce que Csíkszentmihályi (1999) dans son modèle systémique appelle *le champ* (« the field »), constitué par un groupe de personnes ou d'institutions jouant le rôle de *gardiens du domaine* (ou « gatekeepers »). Ce sont ces derniers qui vont accepter ou rejeter la production afin qu'elle puisse ou non faire partie du *domaine* (« the domain »), constitué par un ensemble de connaissances et de symboles culturels. Il semble donc, qu'afin d'obtenir un jugement positif ou une approbation pour une production créative, son auteur doive se familiariser avec les exigences et les valeurs des juges qui vont l'évaluer, afin de les satisfaire.

1.1.2 Evaluation de la créativité en design

Afin d'évaluer si une production est à la fois créative et adaptée, le designer crée ou prend en compte des paramètres pour cette évaluation, dits « référents évaluatifs » (RE), qui peuvent être basés sur des contraintes définies durant la phase de formulation du problème de conception ainsi que sur ses préférences individuelles (Bonnardel, 1996 ; Bonnardel, 2000 ; Bonnardel, 2006 ; Visser, 2009a). L'évaluation peut reposer alors sur l'attribution d'un certain degré d'importance à chaque RE s'appliquant à l'objet en cours de conception. Les problématiques surgissant au cours de l'évaluation sont le plus souvent liées : (1) à un manque de connaissances en design, pouvant être dû à une expérience limitée dans le domaine et à une prise en compte de critères trop restreints, (2) à l'adoption de différents points de vue qui peuvent modifier les buts à atteindre, (3) à la découverte de points faibles dans l'objet en cours de conception qui doivent alors être résolus (Bonnardel, 2006).

Des méthodes existent pour permettre l'évaluation de productions créatives non seulement au cours de son propre travail, mais aussi pour évaluer les productions des autres. Certaines études suggèrent que les mesures portant sur le niveau de créativité des productions sont les plus pertinentes, car elles détermineraient la créativité du design à 45,85%, par comparaison aux mesures de la créativité des individus (19,54%) et du processus lui-même (14,46%) (Demirkan & Hasirci, 2009). La « Technique d'Évaluation Consensuelle » (TEC) proposée par Teresa Amabile (1982), constitue une mesure très répandue et qui a été validée à plusieurs reprises dans le domaine de recherche portant sur la créativité. Elle repose sur l'idée qu'il est possible de mesurer la créativité d'un artefact en prenant compte des évaluations émises de manière indépendante par des experts dans le domaine considéré. Néanmoins, même si cette méthode a été utilisée avec beaucoup de succès dans des différents domaines, elle semble être nettement moins utilisée dans le cas du design, comme le montrent les travaux de Jeffries (2012). De plus, même s'il est important de parvenir à des mesures ayant un bon niveau de fidélité et de validité, la TEC ne permet pas d'explorer les différences entre les opinions exprimées par des juges ayant des points de vues différents, ni l'importance qu'ils attachent à différents critères, en vue d'exprimer leur jugement sur la créativité. Cependant, il nous paraît indispensable de déterminer *pourquoi* une évaluation est positive ou négative, afin d'enrichir, d'un point de vue qualitatif, les connaissances relatives à l'évaluation de la créativité dans le domaine du design.

1.1.3 Critères d'évaluation de la créativité en design

En ce qui concerne les critères importants pour évaluer les productions relevant du design, les avis sont partagés entre différents auteurs spécialisés dans le design. Ainsi, Cropley, Kaufman, et Cropley (2011) défendent surtout l'importance de la « fonctionnalité » (du caractère fonctionnel) des productions, alors que selon Christiaans (2002), « l'originalité » constituerait une valeur ajoutée qui serait déterminante de l'évaluation d'un produit design. Des auteurs parlent du « choc de reconnaissance » (Cropley, Kaufman & Cropley, 2011) et

d'autres évoquent des réactions émotionnelles telles que la surprise, la satisfaction ou la stimulation (Jackson & Messick, 1965). D'autres auteurs encore soulignent l'importance du fait qu'un produit design puisse répondre à toutes les contraintes présentées dans le cahier des charges, car c'est ce qui le différencierait d'autres productions qui, tout en restant créatives, n'auraient aucune fonction pratique. Ont également été proposées des listes de critères qu'un produit devrait remplir afin d'être objectivement considéré comme créatif, comme *The Creative Product Semantic Scale*, développé par Besemer et O'Quin (1987), le *Product Creativity Measurement Instrument* proposé par Horn et Salvendy (2008), ou le *Revised Creative Solution Diagnostic Scale* de Cropley, Kaufman et Cropley (2011). En analysant les travaux de ces différents auteurs (cf. tableau 1) nous avons synthétisé les critères d'évaluation de créativité et obtenu la liste suivante : **Originalité, Caractère adapté au public, Caractère adapté au cahier des charges, Elaboration, Emotions, Esthétique**. Dans ce travail de thèse, nous allons utiliser cette liste de critères pour catégoriser les référents évaluatifs utilisés par les juges qui participent à nos études.

1.1.4 Les profils des juges

Nous nous intéressons aux critères utilisés durant les jugements de créativité en design et cela surtout lors d'évaluations effectuées par des juges ayant des points de vue différents. Ces « points de vue », sont alors supposés dépendre de l'expérience professionnelle passée qui peut influencer la perception que les juges ont du design et de la créativité dans ce domaine et de leur niveau d'expertise dans le domaine du design (mesuré en nombre d'années passées à exercer une activité dans ce domaine). Dans cette thèse, nous souhaitons étudier leurs représentations mentales de la créativité, pouvant être définies, selon Richard (1995), comme des constructions circonstanciées, élaborées dans un contexte particulier et à des fins spécifiques. Selon Johnson-Laird (1995), l'être humain se construit ses représentations internes de la réalité, en percevant ou en imaginant un état des choses. En nous basant sur les travaux d'Herbert Simon (1956), sur la rationalité limitée, et sur ceux de Gigerenzer (2008) portant sur l'utilisation des heuristiques, nous défendons l'idée que les juges utilisent des « raccourcis cognitifs » afin de rendre leurs évaluations moins coûteuses, en prenant comme base de référence des exemples qui sont les plus accessibles en mémoire. Dans le cas de l'évaluation de créativité en design, les juges vont donc utiliser des critères rendus accessibles compte-tenu de leurs expériences professionnelles passées.

Norman (1988) a mis en évidence des différences dans la perception d'un même produit par, d'une part, le concepteur et, d'autre part, l'utilisateur. Glăveanu (2010) a également soutenu que l'évaluation de la créativité dépend du contexte social et culturel du juge - comme son expérience professionnelle ou bien sa position dans la société. Nous supposons que les juges ayant des expériences passées différentes, vont appliquer des heuristiques formées sur la base de ces expériences et, de ce fait, utiliser de manière différente des critères de créativité. Plus précisément, nous allons comparer les jugements des concepteurs et des utilisateurs - cela dans la lignée des travaux de Don Norman, mais nous souhaitons aussi déterminer quelles différences peuvent exister entre des experts en design ayant des profils différents. Dans ce but, avec l'aide d'experts dans le domaine du design, nous avons défini des catégories des juges basées sur la nature de leurs expériences professionnelles : les **designers**, les **directeurs artistiques** et les **enseignants** en design.

Nous établissons également une distinction entre différents niveaux d'expertise reposant sur le nombre d'années d'expérience dans le domaine de design. Ainsi, en nous basant sur de nombreux travaux (Bloom, 1985; Gardener, 1993; Hayes, 1989; Kaufman, 2007) suggérant qu'il faut dix années d'activité intense dans un domaine afin de devenir un expert dans ce domaine, nous avons proposé une seconde catégorisation de nos participants

faisant référence à leur niveau d'expertise : les **experts affirmés** (ayant plus de 10 années d'expérience en design) et les **experts intermédiaires** (ayant entre 5 et 10 ans d'expérience dans ce domaine).

Le débat existe entre les auteurs, concernant la nécessité (ou non) d'utiliser uniquement des experts pour procéder à une évaluation consensuelle des productions créatives. Certaines recherches ont démontré que l'accord inter-juge était plus élevé entre des juges experts qu'entre des non-experts (Kaufman et al., 2008; Lee, Lee, & Young, 2005; Plucker, Holden, & Neustadter, 2008). Néanmoins l'inverse a aussi été démontré (Dollinger, Urban, & James, 2004; Hickey, 2001; Kaufman, Gentile, & Baer, 2005). Selon Caroff et Besançon (2008) les résultats indiquant de meilleurs accords chez des juges non-experts pourraient s'expliquer par le fait que les juges experts développent une compréhension *individuelle* de la créativité au cours de leur expérience professionnelle. De plus, les études dans le champ de l'ergonomie cognitive considèrent qu'il est très important de prendre en compte non seulement les avis d'experts, mais aussi ceux des utilisateurs, lors de l'évaluation des productions qui leur sont destinées. Dans cette perspective, nous avons constitué un groupe, qui rend compte à la fois du **public** ciblé (dont l'opinion doit être prise en compte en tant que futurs utilisateurs des produits design), et qui sert aussi, d'une certaine façon de groupe contrôle (dans le cadre de la seconde catégorisation des participants) car il s'agit aussi de **non experts** en design.

1.2 Partie 2 – Cadre de la recherche et trois études

Dans le cadre de ce travail de thèse, nous avons employé à la fois (1) des méthodes qualitatives, utilisées en ergonomie et en ethnographie afin de collecter des données empiriques, et (2) des méthodes expérimentales, utilisées en psychologie cognitive et se prêtant davantage à des analyses statistiques.

Nous nous sommes centrés sur trois aspects pouvant influencer les jugements de créativité en design :

- **Les critères d'évaluation.** Notre but est de comparer les données empiriques avec les résultats d'études antérieures et de réaliser des études dans un champ spécifique du design, celui du design graphique (moins étudié que celui du design produit ou architectural, et pourtant ayant un grand impact sur la société actuelle). Nous utilisons le terme **réfèrent évaluatif** pour désigner les paramètres identifiés dans les verbalisations des participants et qui servent à évaluer le design. Dans le cadre de ce travail, le terme de **critère** sert de catégorie pour regrouper plusieurs référents évaluatifs. Nous prenons soin de procéder à la catégorisation des données « du terrain » tout en prenant en compte les concepts et les termes utilisés dans la littérature portant sur l'évaluation de créativité en design.
- **Les profils des juges.** Le choix de ces profils a été décrit dans la partie théorique. Notre but est de comprendre l'influence des caractéristiques individuelles des juges sur leurs évaluations de la créativité en design. Nous nous intéressons aux deux caractéristiques suivantes :
 - Leur point de vue professionnel ;
 - Leur niveau d'expertise en design.
- **Le contexte d'évaluation.** Notre travail vise à étudier l'évaluation de créativité en design dans des contextes différents. L'étude de différences individuelles doit, en principe, reposer sur des données issues d'échantillons importants, qu'il est difficile d'atteindre lorsque nous nous intéressons à des profils professionnels spécifiques comme ceux mentionnés plus haut. Nous cherchons, d'une certaine manière, à compenser le nombre limité des

participants en faisant varier les méthodes que nous utilisons auprès d'échantillons plus petits, et tout en favorisant la validité écologique des données qui sont traitées. Nous avons mis en place trois types de situations portant sur des critères différents :

- **Des critères considérés comme importants** pour la créativité, et énoncés par les participants **en dehors de toute situation d'évaluation**, afin d'en inférer des représentations mentales concernant la créativité.
- **Des critères imposés** par l'expérimentateur, sur la base desquels les participants doivent attribuer des scores à des exemples réels de productions design. Nous nous intéressons alors aux liens entre les scores attribués en fonction de ces critères et les scores plus généraux attribués à la « créativité globale », ainsi qu'au caractère consensuel des jugements.
- **Des critères mentionnés de manière spontanée** dans une situation d'évaluation d'exemples de productions design.

1.2.1 Hypothèses

Nous nous attendons à ce que les référents évaluatifs et les critères utilisés pour évaluer la créativité dans le domaine du design graphique diffèrent des ceux utilisés dans des études (plus nombreuses) portant sur le design produit. Le but est ainsi d'approfondir et de compléter les connaissances relatives à l'utilisation des critères dans ce champ du design.

Nous nous attendons également à ce que le point de vue professionnel des juges ainsi que leur niveau d'expertise en design influencent l'utilisation des critères durant l'évaluation de créativité, ainsi que le degré d'accord inter-juges.

Sur ces différentes bases, différents contextes d'évaluation ont été mis en place afin d'identifier les critères utilisés par des groupes des juges, tout en prenant en compte leurs représentations mentales, les liens entre les scores attribués aux critères spécifiques et à la créativité globale, ainsi que les critères exprimés spontanément durant l'évaluation d'exemples de productions design. Les résultats ainsi obtenus constitueront des indicateurs des caractéristiques qui influencent la manière d'évaluer la créativité en design.

1.2.2 Première étude : Critères déclarés comme important pour évaluer la créativité en design.

Le but de cette étude est d'identifier et de recueillir des critères qui reflètent les représentations mentales et les points de vue de nos participants quant à la créativité en design graphique.

Nous avons ainsi poursuivi trois objectifs :

- Premièrement, nous souhaitons savoir quels sont les critères associés à la créativité dans le champ du design graphique. Dans ce but, nous prenons en compte les critères qui sont les plus cités par l'ensemble des participants constitutifs de notre échantillon.
- Deuxièmement, nous voulons déterminer si ces critères diffèrent de ceux proposés dans le champ du design produit.
- Troisièmement, nous souhaitons mettre en évidence les points communs et les différences entre les critères proposés par des participants ayant des profils différents.

L'échantillon pour cette étude est constitué de 43 participants exerçant un métier dans le domaine du design graphique et de 20 participants exerçant un métier dans un autre domaine que le design. Selon leur « point de vue professionnel », nous avons analysé les réponses de 16 designers, 17 directeurs artistiques et 10 enseignants en design. En fonction

du niveau d'expertise, l'échantillon contenait 21 experts affirmés et 22 experts intermédiaires. De plus, nous avons constitué un groupe de 20 personnes représentant le public visé des produits design et en même temps le groupe des personnes non-expertes en design.

Afin de recueillir les données, nous avons mis en place un **questionnaire en ligne**, contenant des questions ouvertes, et nous avons demandé aux participants d'écrire, avec leurs propres mots, les critères qui leur permettent d'évaluer la créativité (1) dans le champ du design graphique et (2) dans le champ du design produit.

Pour procéder à l'**analyse des résultats**, nous nous sommes inspirés de la méthode dite de la *théorie ancrée*, utilisée dans les recherches ethnographiques, qui vise à construire une théorie à partir des données recueillies sur le terrain (à l'opposé de la méthode expérimentale où il s'agit de valider ou non, sur la base des données recueillies, une hypothèse déjà existante ; Glaser & Strauss, 1967). Les réponses des participants ont été listées, puis celles qui pouvaient être considérées comme synonymes ont été regroupées et catégorisées sous les mêmes intitulés de *référents évaluatifs* (RE). Finalement, certains RE ont été regroupés afin de constituer des *critères*, qui ont été définis tout en prenant compte des critères cités dans l'état de l'art (Originalité, Caractère adapté au public, Caractère adapté au cahier des charges, Elaboration, Emotions, Esthétique).

Afin de déterminer quels critères sont cités par le plus grand nombre de participants, nous avons défini des seuils. Si un RE/critère est mentionné par plus de 50% des participants, nous le considérons comme ayant une grande importance pour l'évaluation de la créativité. S'il est mentionné par entre 30% et 50% des participants, nous le considérons comme ayant une importance modérée.

Les résultats obtenus sont les suivants :

- L'Originalité, l'Esthétique et le Caractère adapté au cahier des charges apparaissent comme les critères qui sont les plus associés à la créativité dans le champ design graphique, car ils ont été cités par plus de 50% des participants constitutifs de notre échantillon. Le Caractère adapté au public semble avoir une importance modérée (ce critère étant cité par entre 30 et 50% de l'échantillon)
- Les critères cités comme importants pour la créativité ne sont pas les mêmes dans le champ du design produit (DP) et dans celui du design graphique (DG). Le Caractère adapté au public semble être le plus important pour le DP (critère cité par 73% de l'échantillon), alors que dans le cas du DG, c'est le critère d'Originalité qui est cité en premier lieu (par 71% de l'échantillon). L'Esthétique est prise en compte dans les deux champs du design (par 54% et 63%), de même que l'Originalité (par 71% et 59%). Cependant, le Caractère adapté au cahier des charges semble être considéré comme important uniquement pour apprécier la créativité dans le champ du DG, alors que dans le champ du DP il est mentionné par moins de 30% de l'échantillon). D'autre part, le Caractère adapté au public qui a été mentionné le plus souvent dans le champ du DP, n'a été cité que par 41% de l'échantillon dans le champ du DG.
- Nous avons également identifié, dans le domaine du DG, des différences quant à l'importance accordée aux critères de créativité, entre les participants ayant des points de vue professionnels différents.
 - o Les **designers** semblent avoir des avis communs sur le plus grand nombre de critères (les six critères ont été cités par au moins 30% de ce groupe de participants).

- Les résultats des **directeurs artistiques** se rapprochent le plus de ceux de l'échantillon général.
- Les **enseignants** ont partagé des avis communs sur le nombre le plus bas des critères.

En comparant les résultats des personnes ayant différents niveaux d'expertise en design, nous avons également constaté des différences.

- Les professionnels du design (**experts affirmés et experts intermédiaires**) semblent attacher plus d'importance au Caractère adapté au public et au Caractère adapté au cahier des charges que les non-experts.
- **Les non-experts** présentent moins d'homogénéité, le nombre des RE cités par plus de 30% étant moins élevé que chez les experts.

1.2.3 Deuxième étude : Influence des profils des juges sur la manière d'utiliser des critères de créativité en situation d'évaluation de productions design.

Cette étude est composée des deux parties : la première concerne l'effet du point de vue professionnel, et la deuxième, l'effet de niveau d'expertise des juges sur leurs évaluations. Elle vise à analyser les évaluations de créativité dans le domaine du design graphique dans une situation où des exemples réels de productions design sont présentés et où les juges disposent de critères d'évaluation imposés par l'expérimentateur.

La première partie de l'étude concerne l'effet du point de vue professionnel des juges sur les évaluations.

L'échantillon comprend 20 participants, répartis en 4 groupes : designers, enseignants, directeurs artistiques et audience (correspondant au public visé).

La procédure consistait à leur présenter 21 affiches produites par des étudiants en design, accompagnées du cahier des charges utilisé par les étudiants durant la tâche de conception des affiches. Tous les juges devaient attribuer des scores allant de 1 à 7 en fonction de 4 critères spécifiques : Originalité, Caractère adapté au public, Caractère adapté au cahier des charges et Esthétique, et en fonction de Créativité globale.

Par souci de concision, nous présentons ci-dessous les hypothèses avec les résultats obtenus dans cette étude.

Première hypothèse générale : le point de vue professionnel du juge influence sa manière d'évaluer la créativité en design graphique. Cette hypothèse a été **validée**, car les résultats d'ANOVA des analyses statistiques (ANOVA) ont révélé des différences significatives entre les scores attribués par les juges ayant différents points de vue professionnels.

Deuxième hypothèse générale : le point de vue professionnel des juges a une influence sur les variations de l'accord inter-juges lors des scores attribués sur la base de critères spécifiques de créativité imposés par l'expérimentateur. Cette hypothèse a été **validée** par les résultats d'analyses statistiques (alpha de Cronbach). Selon ces résultats, l'indice d'accord inter-juges est élevé pour des critères différents, selon le groupe des juges.

Troisième hypothèse générale : les critères spécifiques contribuent de manière différente à l'évaluation de la créativité globale, selon le point de vue professionnel des juges. Cette hypothèse a été **validée**, car l'analyse des régressions a montré que, selon le groupe des juges, des critères spécifiques différents ont eu une influence significative sur l'évaluation globale de la créativité.

Sur la base de ces résultats, nous avons mis en évidence, que le point de vue professionnel influence effectivement la manière dont les juges procèdent à des évaluations de créativité dans le domaine du design graphique. Pour en savoir plus sur la nature de ces jugements, nous avons émis et testé des **hypothèses spécifiques** :

H1. Dans le domaine du design, de nombreux auteurs préfèrent imposer aux participants des listes de critères spécifiques plutôt que de demander aux juges une appréciation globale de la créativité. Si cette méthode est plus adaptée au design graphique, que les méthodes portant uniquement sur les jugements de la créativité globale (par exemple TEC) nous nous attendons à ce que l'accord inter-juge soit plus élevé pour les critères spécifiques que pour la créativité globale.

Cette hypothèse a été validée de manière partielle, car les résultats des analyses statistiques (alpha de Cronbach) ont montré, dans tous les groupes de juges, un accord inter-juges, qui était plus élevé pour au moins un critère spécifique à celui concernant la créativité globale. Cependant, ce constat n'a pas pu être fait pour tous les critères spécifiques. Néanmoins, les résultats montrent que certains groupes de juges développent une approche particulière envers le design graphique, comme s'ils avaient plus des connaissances partagées sur certains critères spécifiques que sur d'autres.

H2. Suite au modèle de Norman (1988) et aux résultats de l'étude pilote que nous avons réalisée (Wojtczuk & Bonnardel, 2012), nous nous attendons à ce que les critères spécifiques qui guident le plus l'évaluation de la créativité globale par les directeurs artistiques soient partagés avec les critères pris en compte par le public visé (les directeurs artistiques étant censés connaître précisément les préférences du public), alors que les enseignants devraient partager davantage les critères des designers (puisque professionnellement, ce sont eux qui ont en charge le transfert aux élèves de connaissances utilisées par les designers professionnels).

Cette hypothèse a été validée : pour les designers et les enseignants, les scores qu'ils attribuent à l'Originalité et au Caractère adapté au cahier des charges semblent affecter le plus les scores attribués à la Créativité globale. Il en est de même pour les directeurs artistiques et le public pour lesquels le caractère adapté au cahier des charges semble contribuer le plus à l'évaluation de la Créativité globale. L'interprétation de ces résultats est la suivante : il existerait une distinction entre des profils « créatifs » et des profils « centrés utilisateur ». Selon la nature des expériences professionnelles antérieures, un profil ou un autre serait plus particulièrement favorisé chez chaque groupe de juges (même si ces profils ne sont pas pour autant exclusifs) ce qui aurait un effet sur l'accessibilité des heuristiques d'évaluation et de certains critères.

H3. En accord avec Hooke, Nakamura et Csikszentmihalyi (2003) les directeurs artistiques et les enseignants, que nous considérons comme les *gardiens du domaine*, devraient avoir un niveau élevé de connaissances communes relatives au design graphique. Ceci leur permettrait de partager leurs avis sur le degré d'innovation des productions design et sur leur caractère adapté aux différentes exigences. Aussi, nous nous attendons à ce que leur niveau d'accord inter-juges soit plus élevé que celui des autres groupes de participants.

Cette hypothèse n'a pas pu être validée. Nous n'avons pas pu observer chez les enseignants et les directeurs artistiques, un accord inter-juges plus élevé que chez les autres participants. Il semble donc, que le point de vue de ceux qui peuvent être considérés comme les gardiens de domaine n'est pas l'objet de davantage d'homogénéité que celui des autres juges. Il se peut que l'expérience individuelle acquise au cours des années ait eu un impact important sur tous les juges, quel que soit leur profil.

H4. Nous nous attendons également à ce que les scores attribués pour la Créativité globale par les designers soient basés surtout sur l'Originalité et sur le Caractère adapté au public. Cela, car leur manière de percevoir la créativité est influencée par les caractéristiques de l'activité proprement dite de conception qui amène les designers à, d'une part, rechercher des idées nouvelles (Bonnardel, 2000, 2006) et d'autre part prendre en compte différents points de vue (Bonnardel & Summer, 1996). En outre, dans la lignée des travaux de Lera (1981), nous nous attendons à ce que les designers attribuent une importance variable et hétérogène aux contraintes du projet et, de ce fait, obtiennent un accord inter-juges bas en ce qui concerne le Caractère adapté au cahier des charges.

Cette hypothèse a été validée. Dans le groupe des designers, les scores attribués à l'Originalité et au Caractère adapté au cahier des charges ont influencé de manière hautement significative les scores attribués à la Créativité globale. Le contraire a été observé pour le Caractère adapté au cahier des charges, dont le score ne semble pas lié au score de Créativité globale. Cette dernière observation s'est retrouvée aussi pour d'autres groupes de juges. Il se peut donc que le problème envisagé d'attribution d'importance aux contraintes mentionnées dans le cahier des charges soit généralisable à tous les profils de participants.

H5. En ce qui concerne le public, en nous basant sur l'hypothèse des théories implicites (Chan & Chan, 1999) nous nous attendons à ce que les évaluations effectuées par ce groupe soient basées sur leurs représentations ou croyances personnelles et non sur des théories partagées de manière explicite. Aussi, leurs évaluations de la Créativité globale devraient être moins influencées par les critères spécifiques et un accord inter-juges plutôt bas devrait être constaté.

Cette hypothèse a été partiellement validée. Aucun critère spécifique n'a influencé de manière significative les évaluations de la Créativité globale effectuées par ce groupe, ce qui suggère que leurs représentations concernant le domaine du design ne sont pas partagées, probablement en raison d'un manque des connaissances dans ce domaine. Néanmoins, leur accord inter-juges pour des critères spécifiques, comparé à d'autres participants, est l'un des plus élevés, ce qui va en faveur de considérations selon lesquelles des juges non-experts peuvent fournir des évaluations consensuelles portant sur la créativité (Dollinger, Urban, & James, 2004; Hickey, 2001; Kaufman, Gentile, & Baer, 2005).

La seconde partie de l'étude concerne l'effet du niveau d'expertise des juges sur les évaluations.

L'échantillon comprend 21 participants, répartis en 3 groupes : experts affirmés en design, experts intermédiaires et non-experts.

La procédure est la même que dans la première partie de l'étude et les mêmes affiches ont été soumises à l'évaluation.

Première hypothèse générale : le niveau d'expertise du juge influence sa manière d'évaluer la créativité dans le champ du design graphique. Cette hypothèse a été **validée**, car les résultats des analyses statistiques (ANOVA) ont révélé des différences significatives entre les scores attribués par les juges ayant différents niveaux d'expertise en design.

Deuxième hypothèse générale : le niveau d'expertise des juges influence le degré d'accord inter-juges lors des scores attribués sur la base de critères spécifiques. Cette

hypothèse a été **validée** par les résultats des analyses statistiques (alpha de Cronbach). Selon ces résultats, l'accord inter-juges s'est révélé élevé pour des critères différents, selon le groupe des juges.

Troisième hypothèse générale : les critères spécifiques contribuent de manière différente à l'évaluation de la Créativité globale, selon le niveau d'expertise des juges. Cette hypothèse a été validée, car l'analyse des régressions a montré que, selon le groupe des juges, des critères spécifiques différents ont eu une influence significative sur la Créativité globale.

Hypothèses spécifiques :

H1. De même que dans la première partie de l'étude, nous nous attendons à ce que l'accord inter-juges soit plus élevé pour les critères spécifiques que pour la Créativité globale.

Cette hypothèse a été validée de manière partielle. Dans tous les cas, l'accord inter-juge a été plus élevé pour un ou plusieurs critères spécifiques, que pour la Créativité globale. Cependant, nous avons constaté certains critères spécifiques, pour lesquels cet accord était bien plus bas que pour la Créativité globale.

H2. En accord avec les résultats de Kaufman et al. (2008), Lee, Lee et Young (2005) et ceux de Plucker, Holden et Neustadter (2008), l'accord inter-juges devrait être moins élevé pour des juges non-experts.

Cette hypothèse n'a pas pu être validée (les coefficients d'alpha de Cronbach ne sont pas moins élevés pour les non-experts, par rapport à d'autres groupes des juges). Ces résultats rejoignent ceux des auteurs qui soutiennent que les juges non-experts peuvent fournir des évaluations consensuelles aussi bien que les experts (Dollinger, Urban, & James, 2004; Hickey, 2001; Kaufman, Gentile, & Baer, 2005). Cependant, les évaluations portées par des juges non-experts semblent être consensuelles pour d'autres critères que celles portées par d'autres groupes. Alors qu'on trouve un accord entre les juges non-experts en ce qui concerne l'attribution de scores lors de l'évaluation du Caractère adapté au public, chez les autres juges un tel accord s'observe pour le critère d'Originalité.

H3. Compte-tenu de la « règle » supposée des 10 années nécessaires pour obtenir une expertise dans un domaine (Bloom, 1985 ; Gardener, 1993 ; Hayes, 1989 ; Kaufman, 2007), nous nous attendons à des différences de degré d'accord inter-juges, entre les experts affirmés et les experts intermédiaires.

Cette hypothèse n'a pu être validée, puisque les deux groupes de juges semble porter des jugements consensuels sur le même critère (l'Originalité) et leur niveau d'accord ne semble pas diverger. Ainsi, le fait de bénéficier de plus de 10 ans d'expérience n'impacterait ni la nature, ni le niveau d'accord inter-juges des experts.

1.2.4 Troisième étude : Critères spontanément utilisés en situations d'évaluation de la créativité dans le champ du design graphique.

Cette troisième étude a été réalisée dans trois buts :

- Identifier les RE et les critères les plus utilisés durant des verbalisations spontanées accompagnant des évaluations libres de la créativité en design.
- Comparer les RE utilisés lors d'une évaluation réflexive et lors d'une évaluation viscérale (niveaux de perception du design décrits par Don Norman en 2004) de la créativité dans le domaine du design, avec les RE utilisés lors d'une évaluation basée sur des préférences personnelles.

- Analyser les différences et les points communs entre les RE et les critères utilisés par les juges ayant des points de vue professionnels et des niveaux d'expertise différents.

L'**échantillon** comprend 23 participants. En ce qui concerne le point de vue professionnel, nous avons pris en compte les réponses des 6 designers, 4 directeurs artistiques et 6 enseignants en design. En ce qui concerne le niveau d'expertise en design, l'échantillon comprend 8 experts affirmés, 8 experts intermédiaires et 7 personnes représentant le public potentiel des produits design et correspondant, en même temps, au groupe des juges non-experts.

Pour recueillir les données, nous avons conduit des **entretiens semi-directifs**, qui consistaient à demander aux participants d'évaluer la créativité de 8 affiches produites par les étudiants en design. Ces entretiens ont été réalisés en fonction de 4 étapes : (1) choisir les affiches qui, au premier abord sont jugées comme les plus créatives (évaluation au niveau viscéral), puis justifier ces choix ; (2) évaluer à voix haute la créativité de chaque affiche (évaluation au niveau réflexif) ; (3) choisir les affiches les plus créatives, en fonction de l'évaluation réflexive mise en œuvre lors de l'étape précédente ; (4) choisir des affiches basées en fonction des préférences personnelles des juges, puis justifier ces choix.

L'**analyse des résultats**, de même que dans la première étude de ce travail de thèse, a été inspirée de la méthode dite de *théorie ancrée* et elle a consisté à identifier, dans les verbalisations des juges, les référents évaluatifs (RE) puis à les catégoriser sous forme des critères.

Les résultats obtenus sont les suivants :

- 1) Le critère d'**Esthétique** a été le plus cité (43% de tous les RE sont liés à ce critère). Cependant ce résultat semble, en grande partie, dû à la haute granularité de ce critère. Par exemple, les juges peuvent consacrer beaucoup de temps à analyser l'utilisation de polices de caractères ou de couleurs. Il semble que la créativité des éléments liés à l'Esthétique puisse être jugée indépendamment des autres aspects des productions dans le domaine du design, à la fois en fonction de son caractère innovant et de son caractère adapté.
- 2) La catégorisation des RE a montré que, lors de l'expression spontanée d'évaluations de la créativité, les RE relevant des catégories « Caractère adapté au cahier des charges » et « Caractère adapté au public » sont souvent confondus. Nous pouvons les regrouper sous le même intitulé : « **Caractère adapté au contexte** » (qui représente alors 23% des RE).
- 3) Le critère d'**Originalité** est moins cité que ce à quoi nous aurions pu nous attendre à l'issue des résultats de l'étude 1 (uniquement 18% des RE, comparé à 68% durant la première étude). Il se peut qu'il soit plus facile de se centrer sur des éléments concrets, comme les éléments de design relevant de l'Esthétique, que sur des éléments plus abstraits, comme ceux relevant de l'Originalité.
- 4) Les **Emotions** (qui représentent 16% des RE) prennent une place importante et, suite aux analyses effectuées, font partie des 4 critères les plus utilisés dans le cadre de verbalisations spontanées lors de l'évaluation d'exemples réels de productions relevant du design graphique, alors qu'elles étaient faiblement citées durant la première étude liées aux représentations mentales des participants.

En ce qui concerne les **choix résultant d'une évaluation viscérale**, émis par les juges, les 4 critères considérés sont repartis de manière assez équilibrée. Cependant, l'Originalité est le RE qui est exprimé le plus souvent pour justifier ces choix « viscéraux ». Durant l'**évaluation réflexive**, les juges semblent se concentrer surtout sur l'Esthétiques des

productions, ainsi que sur leur Caractère adapté au contexte. Enfin, lorsqu'il s'agit d'une **évaluation basée sur les critères personnels** des juges, il semble que ces derniers font surtout référence à l'Esthétique et aux Emotions.

- L'analyse des écarts réduits met en évidence les différences les plus marquées entre les fréquences d'utilisation des critères par les **juges ayant des points de vue différents**. Nous remarquons que le critère d'Esthétique est surtout mentionné par les designers et les directeurs artistiques, alors que le public lui attache relativement peu d'importance. Cela semble être le contraire pour les Emotions, qui sont souvent mentionnées par le public, mais peu par les juges considérés comme constituant des gardiens du domaine (enseignants et directeurs artistiques). Ces derniers, en revanche, semblent accorder davantage d'attention à l'Originalité des productions.
- De façon similaire aux résultats obtenus lors de la deuxième étude présentée dans cette thèse, les jugements portant sur le Caractère adapté au contexte divergent (un même élément pris en compte pouvant être jugé comme « adapté » par certains juges et comme « non adapté » par d'autres). De telles divergences s'observent surtout lorsqu'il s'agit de l'utilisation de symboles (visuels ou choix de couleurs) ayant pour but de réaliser une analogie avec la thématique de l'affiche, qui amènent les juges à faire des interprétations différentes. Ces interprétations, portant sur la compréhensibilité des symboles et leur fréquence d'usage dans le monde du design, influencent aussi bien leurs jugements portant sur le Caractère adapté que sur l'Originalité.

1.3 Partie 3 – Discussion

Les résultats des trois études présentées ont permis de déterminer et de préciser les critères qui sont utilisés lors d'évaluations dans le champ du design graphique. De plus, ils nous ont permis de comprendre l'influence que les différents profils de juges peuvent avoir sur la manière d'évaluer la créativité tout en utilisant ces critères.

Trois méthodes différentes ont été mises en œuvre pour tester des hypothèses générales (présentées dans la partie 2 de cette thèse).

Dans un premier temps, nous avons vérifié que des différences existaient lors de l'évaluation de productions relevant du champ du design graphique et de celui du design produit. Un questionnaire ouvert nous a permis d'inférer les représentations mentales concernant la créativité en design graphique et de rendre compte des différences existant dans ces deux champs du domaine de design. Il semble que la créativité dans le champ du design graphique dépende davantage de l'originalité des idées et du respect du cahier des charges, alors que dans le champ du design produit, la créativité est surtout associée au caractère adapté à l'utilisateur.

Le but de cette thèse était, également, de compléter et d'approfondir les connaissances quant à l'utilisation de critères dans le champ du design graphique. Les études qui ont été réalisées ont montré que le critère d'**Originalité**, déclaré comme important par le plus grand nombre de personnes ayant répondu au questionnaire (étude 1), n'est pas forcément le plus utilisé en situations d'évaluation libre d'exemples réels de productions design. Dans ces situations, il semble que l'Originalité soit surtout citée durant les évaluations reposant sur une approche viscérale. L'importance que ce critère prend, au niveau des représentations mentales et lors d'une évaluation viscérale, pourrait indiquer que, juger de

la créativité dans le champ du design graphique, en se basant surtout sur le critère d'originalité, correspond à une application de l'heuristique la plus accessible chez une majorité de juges (Gigerenzer, 2008). De plus, les juges (surtout des experts intermédiaires et confirmés dans le domaine du design) semblent présenter un haut niveau d'accord quand il s'agit d'attribuer des scores sur ce critère. Les analyses portant sur l'évaluation réflexive indiquent que certains juges experts ont tendance à considérer ce critère, non seulement, comme faisant référence à de l'originalité au niveau de la production ou du produit, mais aussi, au niveau de toute la démarche mise en œuvre par un individu créatif (comme un designer). Cela rend compte du fait que les paramètres de la créativité proposés par Rhodes (1961) interagissent entre eux, même « dans l'esprit » de personnes engagées dans des situations d'évaluation.

L'Esthétique a été l'un des critères les plus souvent cités durant les verbalisations spontanées mais, les résultats de l'étude expérimentale (étude 2) ont indiqué que ce critère semble moins influencer l'évaluation de la Créativité globale que les autres critères spécifiques, sauf les directeurs artistiques – ces derniers semblant établir plus de liens entre Esthétique et Créativité que les autres juges. L'analyse des verbalisations lors des évaluations (étude 3) nous a montré que les éléments esthétiques, comme la couleur, la composition et les polices de caractère, sont souvent analysés à part, chacun de ces éléments méritant une analyse portant à la fois sur son caractère original et sur son caractère adapté. D'une certaine manière, les juges semblent même dissocier la créativité globale d'une œuvre de son esthétique.

Le **Caractère adapté au cahier des charges** est apparu important pour la créativité en design graphique, à la fois au niveau des perceptions et des croyances, mais aussi lors de l'attribution de scores aux exemples réels de productions design. Cependant, les résultats de la seconde étude indiquent que c'est ce critère qui est source de désaccord entre les juges. L'analyse des verbalisations (étude 3) peut fournir des indications possibles quant aux raisons de ce désaccord. D'une part, les juges s'expriment de manière telle qu'il est difficile de dissocier dans leurs avis ce qui concerne le caractère adapté au cahier des charges et ce qui concerne le degré de compréhension potentielle de l'œuvre par le public. D'autre part, les opinions divergent surtout quand il s'agit d'évaluer les métaphores utilisées dans les visuels des affiches, car leur caractère adapté est souvent jugé sur la base de connaissances et de valeurs personnelles du juge. Cela confirme les observations de Lera (1981) suggérant que, lors de l'évaluation du caractère adapté, les juges ont tendance à attribuer des valeurs personnelles, surtout lorsque le cahier des charges n'est pas exhaustif.

Le **Caractère adapté au public** a été, à l'issue de certains de nos résultats, fusionné avec le Caractère adapté au cahier des charges. Cependant, nous pouvons constater que, même si ce critère ne semble pas être le plus fréquemment présenté comme important, au niveau des représentations mentales concernant la créativité dans le champ du design graphique (étude 1), les juges prennent ce critère en considération au moment d'attribuer des scores aux productions (étude 2). Pour certains groupes, le niveau d'accord inter-juges sur ce critère est bien plus significatif que sur celui qui fait référence au cahier des charges. L'étude des verbalisations (étude 3), nous laisse penser que ce critère est davantage lié aux aspects émotionnels que fonctionnels du design, ce qui nous incite à rejoindre les propositions d'auteurs comme Tracinsky, Shoval-Katz et Ikar (2000) ou Norman (2004) qui considèrent que l'aspect émotionnel influence la perception de la fonctionnalité des produits.

Les résultats de cette thèse présentent, à notre avis, un intérêt à la fois dans des perspectives appliquées et fondamentales. D'un point de vue fondamental, les connaissances dans le domaine du design peuvent être enrichies par de données ayant une forte valeur écologique et tenant compte à la fois du contexte d'évaluation et de différents profils de juges. Cela d'autant plus que, le domaine du design graphique n'a été, jusque-là, que peu étudié dans le cadre de recherches portant sur la créativité. Ces informations pourraient être également utiles pour tout type d'étude utilisant la méthode des juges comme moyen de mesure de la créativité. D'un point de vue appliqué, les designers professionnels pourraient bénéficier d'informations concernant la manière dont différents juges approchent et perçoivent le design graphique. Cela leur permettrait, d'anticiper les critiques associés à leur travail et de préparer une argumentation visant à défendre leur vision ou leurs idées de la manière la plus convaincante. Dans le domaine de l'éducation et de la pédagogie du design, connaître le point de vue des professionnels pourrait permettre, d'une part, aux enseignants de mieux adapter les critères d'évaluation des travaux d'étudiants aux exigences du monde du travail et, d'autre part, aux étudiants de s'approprier de tels critères.

TABLE OF CONTENTS

ACKNOWLEDGMENTS	3
RÉSUMÉ	5
MOTS CLÉS	5
ABSTRACT	7
KEYWORDS	7
RÉSUMÉ SUBSTANTIEL	9
1.1 Partie 1 - Etat de l'art.....	10
1.1.1 Créativité	10
1.1.2 Evaluation de la créativité en design.....	11
1.1.3 Critères d'évaluation de la créativité en design.....	11
1.1.4 Les profils des juges.....	12
1.2 Partie 2 – Cadre de la recherche et trois études	13
1.2.1 Hypothèses	14
1.2.2 Première étude : Critères déclarés comme important pour évaluer la créativité en design.....	14
1.2.3 Deuxième étude : Influence des profils des juges sur la manière d'utiliser des critères de créativité en situation d'évaluation de productions design.....	16
1.2.4 Troisième étude : Critères spontanément utilisés en situations d'évaluation de la créativité dans le champ du design graphique.....	19
1.3 Partie 3 – Discussion	21
TABLE OF CONTENTS	25
INDEX OF FIGURES	31
INDEX OF TABLES	33

INTRODUCTION.....	36
1 Studying creativity assessment in design	36
2 The field of graphic design.....	37
3 Aims of the research	38
4 Combining multiple methods	38
5 Thesis structure.....	39
PART I – STATE OF THE ART	40
Chapter 1 – Concepts of <i>creativity</i> and <i>design</i>	40
1 Creativity	40
1.1 Definition of creativity	40
1.2 Perceptions of creativity.....	41
1.3 Taxonomy of creativity: process, person, place and product.....	42
1.4 System-oriented models of creativity.....	43
2 Design.....	45
2.1 Emergence of the cognitive approach to design.....	45
2.2 Specificity of the design process	46
2.3 Stages of the design process.....	47
2.3.1 Formulation and reformulation of the problem	48
2.3.2 Generation of solutions.....	48
2.3.3 Evaluation of solutions	49
Chapter 2 Creativity assessment.....	50
1 Taxonomy of creativity measures	50
2 Assessing creative products.....	51
2.1 Why focus on product creativity?	52
2.2 Existing measures of product creativity	52
2.4 Consensual Assessment Technique (CAT)	53
2.4.1 Theoretical background to the CAT	53
2.4.2 CAT limitations in the context of design.....	54
3 The model of judgment.....	55
3.1 Lens model of judgment.....	55
3.2 Use of heuristics	55
4 Influence of judge characteristics on creative product assessment	56
4.1 Assessing design: differences in mental representations of the same product	56
4.1.1 Designers’ judgments	58
4.1.2 Domain gatekeepers’ judgments	58
4.1.3 Laypeople or users’ judgments.....	59

4.2	Differences in creativity assessments depending on the judges' level of expertise.....	60
5	Criteria for assessing design creativity	61
5.1	Criteria used to assess creative design.....	61
5.1.1	Overview of criteria based on existing studies.....	62
5.1.2	Definitions of chosen criteria	64
6	Different levels of design.....	65
PART II – RESEARCH FRAMEWORK AND THREE STUDIES		68
Chapter 3 Research framework.....		68
1	Studying variability of creativity evaluations	68
2	Research questions and general hypotheses	69
3	Methodologies	73
3.1	Data collection.....	73
3.2	Data analysis.....	74
Chapter 4 – Criteria declared as important for the design creativity assessment		77
1	Method	77
1.1	A survey with open questions.....	77
1.2	Data analysis.....	78
2	Participants.....	79
2.1	Recruitment	79
2.2	Establishing specialization profiles	79
2.3	Establishing experience profiles.....	80
2.4	Final sample.....	80
3	Material.....	81
4	Procedure for criteria finding.....	81
4.1	First step: identifying codes.....	81
4.2	Second step: categorisation and ER finding.....	81
4.3	Third step: identification of criteria in creative assessment	82
5	Categorisation results and Evaluative Referents finding.....	82
5.1	Similarities and differences between mental representations of the GD and PD creativity	84
5.2	ER depending on the participants' professional backgrounds.....	85
5.3	ER depending on the participants' level of experience in design.....	86
6	Identification of criteria in creative assessment.....	88
6.1	Criteria depending on the participants' professional backgrounds	89
6.2	Criteria depending on the participants' level of experience in design	90
7	Discussion.....	92

Chapter 5. The influence of judges profile on the criteria used in the real situation of the creative design assessment	95
1 Pilot study in the field of product design.....	95
1.1 General goals of the experimentation.....	97
2 Study 1: The influence of judges' backgrounds on the graphic design assessment	98
2.1 Hypotheses	98
2.2 Participants	99
2.3 Material	100
2.4 Procedure.....	100
2.5 Results	100
2.5.1 The general impact of judges' backgrounds on the assessments of the design creativity	100
2.5.2 The impact of judges' backgrounds on the degree of the inter-judge agreement	101
2.5.3 Differences within the criteria that contributed to overall creativity assessments, depending on the judges' backgrounds	102
2.6 Discussion	103
3 Study 2: The influence of judges' level of experience in the field of design on the assessment of graphic design	105
3.1 Hypotheses	106
3.2 Participants	107
3.3 Material	107
3.4 Procedure.....	107
3.5 Results	107
3.5.1 The general impact of judges' level of experience within the design field on the assessments of the design creativity	107
3.5.2 The impact of judges' level of experience within design field on the degree of the inter-judge agreement.....	108
3.5.3 Differences in the criteria that contributed to overall creativity assessments, depending on the judges' level of experience	108
3.6 Discussion	109
Chapter 6 – Criteria used spontaneously in real situation of creative design assessment	111
1 Methodology.....	112
1.1 Data analysis	112
2 Participants	113
2.1 Recruitment	113
2.2 Establishing specialization profiles	113
2.3 Establishing experience profiles.....	114

2.4	Final sample.....	114
3	Material.....	114
4	Procedure	115
4.1	Interviews	115
4.2	Analyses	115
5	First stage of the study: categorisation results	116
5.1	Method of analyses.....	116
5.2	First step: identifying codes.....	116
5.3	Second step: categorisation and ER finding.....	116
5.4	Third step: criteria finding.....	117
5.5	Identification of ERs.....	117
5.6	Observations made during verbatim analyses	123
5.7	Criteria-finding.....	124
5.8	Conclusion.....	125
6	Second stage of the study: criteria used by judges for different types of assessments.....	127
6.1	The distribution of criteria used by all judges during different types of assessments.....	127
6.1.1	Global results.....	128
6.1.2	Choices based on the visceral assessment.....	129
6.1.3	Choices based on reflective assessment	129
6.1.4	Choices based on personal preferences	131
	Conclusions	132
6.2	132
7	Third stage of the study: criteria used by judges depending on their backgrounds and their levels of experience in design.....	134
7.1	The distribution of judges' assessments with regard to their different profiles	134
7.1.1	Results depending on judges' backgrounds	134
7.1.2	Results depending on judges' levels of experience.....	135
7.2	The distribution of the criteria and ER use, within the assessments of judges with different profiles	136
7.2.1	Results depending on judges' backgrounds	136
7.2.2	Results depending on judges' levels of expertise.....	139
7.3	Conclusion.....	141
	PART III – DISCUSSION	145
	Chapter 7 General discussion and research perspectives.....	145
1	Summary of our findings	146

1.1	Summary of our findings concerning the criteria and ERs used to assess the creativity of graphic design	146
1.2	Summary of our findings concerning the influence of judges' profiles on creativity assessments of graphic design.....	148
1.2.1	Professional background	148
2	Research perspectives.....	150
3	Perspectives for applied design	152
BIBLIOGRAPHY		155
APPENDICES.....		169
Appendix 1 - Online questionnaire used for the first study.....		169
Appendix 2 - Evaluative referents for Graphic Design (GD) and Product Design (PD)		171
Appendix 3 - Number of evaluative referents cited by judges, depending on their (a) backgrounds and (b) levels of experience in design.....		175
Appendix 4 - Indications for participants of third study		177
Appendix 5 - Posters presented to participants of the third study.....		179
Appendix 6 - Criteria used during the entire duration of the interviews		181

INDEX OF FIGURES

Figure 1: Multicomponent model (Lubart et al., 2003; Sternberg & Lubart 1995)	44
Figure 2: Systemic model (Csíkszentmihályi, 1999).....	44
Figure 3: Design stages: (1) Asimov, 1962; (2) McNeil, Gero, & Warren, 1998 (from Bonnardel, 2009)	47
Figure 4: Creative stages: (1) Wallas, 1927; (2) Gelb, 1996; (3) Amabile, 1996 (from Bonnardel, 2009)	47
Figure 5: The information processing model (Zeng, Salvendy & Zhang, 2009).	60
Figure 6: Percentages of occurrences of criteria used by judges during the entire interview.	128
Figure 7: Percentages of occurrences of criteria on which judges base their visceral assessments.....	129

Figure 8: Percentages of occurrences of criteria resulting with positive statements on which judges based their choices after reflective assessments.	130
Figure 9: Percentages of occurrences of criteria on which judges based their personal choices.....	131
Figure 10: Multiple Correspondence Analysis of the distribution of criteria used during the whole interview, across the judges' backgrounds.....	135
Figure 11: Multiple Correspondence Analysis of the distribution of criteria used during the whole interview, across the judges' levels of experience.	135
Figure 2: Numbers of occurrences for ERs used during the whole interview, within groups of judges with different backgrounds.....	181
Figure 2: Numbers of occurrences for each ER during the first stage of interviews (visceral choices) within groups of judges with different levels of experience in design field.....	182
Figure 2: Numbers of occurrences for ERs used during the first stage of interviews (visceral choices), within groups of judges with different backgrounds.....	183
Figure 2: Numbers of occurrences for ERs used during the first stage of interviews (visceral choices), within groups of judges with different levels of experience.	184
Figure 2: Numbers of occurrences for ERs used during the second stage of interviews (reflective assessments), within groups of judges with different backgrounds.....	185
Figure 2: Numbers of occurrences for ERs used during the second stage of interviews (reflective assessments), within groups of judges with different levels of experience.	186
Figure 2: Numbers of occurrences for ERs used during the third stage of interviews (reflective choices), within groups of judges with different backgrounds.	187
Figure 2: The numbers of occurrences of positive statements used by judges with different levels of professional to describe posters that they chose during the third stage of the interview (reflective choices).	188
Figure 2: Numbers of occurrences for each ER during the fourth stage of interview (personal preferences) within groups of judges with different backgrounds.	189
Figure 2: Numbers of occurrences for each ER during the fourth stage of interview (personal preferences) within groups of judges with different levels of experience.	190

INDEX OF TABLES

Table 1: Criteria to assess creativity studied by different authors, with the proposition of synthesis under common labels	63
Table 2: The percentages of ERs related with creativity that were cited by all the participants in the area of GD and PD	83
Table 3: The most cited Evaluative Referents within the designers group	85
Table 4: The most cited Evaluative Referents within the art directors group	86
Table 5: The most cited Evaluative Referents within the teachers group	86
Table 6: The most cited Evaluative Referents within the experts group	87
Table 7: The most cited Evaluative Referents within the intermediary experts group	87
Table 8: The most cited Evaluative Referents within the laypeople group	87
Table 9: Categorisation of ERs related to GD and PD areas, with linking to the creative design criteria from literature	88
Table 10: The most cited criteria within the whole sample	89
Table 11: The most cited criteria within the designers group	89
Table 12: The most cited criteria within the art directors group	90
Table 13: The most cited criteria within the teachers group	90
Table 14: The most cited criteria within the asserted experts group	91
Table 15: The most cited criteria within the intermediary experts group	91
Table 16: The most cited criteria within the laypeople group	91
Table 17: The distribution of partial correlations between the specific criteria scores and the overall scores, depending on the judges' professional background. The strongest partial correlations for each group of judges are marked in blue	96

Table 18: Distribution of inter-judge agreement (expressed in Cronbach's alpha) within posters assessments based on 5 criteria depending on judges' backgrounds.	101
Table 19: The distribution of significant effects of the specific criteria scores on the overall creativity score, depending on the judges' professional background.....	102
Table 20: Distribution of inter-judge agreement (expressed with Cronbach's alpha) within poster assessments based on 5 criteria depending on judges' level of experience in design	108
Table 21: The distribution of significant effects of the specific criteria scores on the overall creativity score, depending on the judges' level of professional experience in design. ...	109
Table 22: Categorization of ERs with linking to the creative design criteria from literature	125
Table 23: Criteria and ERs used by judges during the entire interview.....	128
Table 24: Criteria and ERs used by judges to make choices based on visceral assessment	129
Table 25: Criteria and ERs used by judges to make choices based on reflective assessment.	130
Table 26: Criteria and ERs used by judges to make choices based on personal preferences.	131
Table 27: Comparison of creativity ERs used for visceral and reflective assessments	133
Table 28: Comparison of creativity ERs used for assessments of creativity and those based on personal preferences	133
Table 29: Criteria most and least used by participants with different backgrounds, for choices based on different types of assessments.	137
Table 30: ERs most and least used by participants with different backgrounds, for choices based on different types of assessments.	138
Table 31: Criteria most and least used by participants with different backgrounds, for choices based on different types of assessments.	139
Table 32: ERs most and least used by designers with different levels of expertise in design for choices based on different types of assessments.	140
Table 30: Categorized ERs mentioned as the most important for creativity in both, graphic and product design areas (during the first study).	171
Table 30: ERs cited during the first study, according to the judges backgrounds.	175
Table 30: ERs cited during the first study, according to the judges levels of experience in design.	176
Table 30: Standardized Deviates for the number of occurrences of ERs, within the groups of judges with different professional backgrounds.	181
Table 30 Standardized Deviates for the number of occurrences of ERs, within the groups of judges with different levels of professional experience.....	182
Table 30 Standardized Deviates for the number of occurrences of ERs used during the first stage of interviews, within the groups of judges with different backgrounds.....	183

Table 30 Standardized Deviates for the number of occurrences of ERs used during the first stage of interviews, within the groups of judges with different levels of professional experience.....	184
Table 30 Standardized Deviates for the number of occurrences of ERs used during the second stage of interviews, within the groups of judges with different backgrounds.....	185
Table 30 Standardized Deviates for the number of occurrences of ERs used during the second stage of interviews, within the groups of judges with different levels of professional experience.....	186
Table 30 Standardized Deviates for the number of occurrences of ERs with positive statements to describe posters chosen during the third stage of the interview (reflective choices), within the groups of judges with different backgrounds.....	187
Table 30 Standardized Deviates for the number of occurrences of ERs with positive statements to describe posters chosen during the third stage of the interview (reflective choices), within the groups of judges with different backgrounds.....	188
Table 30 Standardized Deviates for the number of occurrences of criteria with positive statements to describe posters chosen during the fourth stage of the interview (personal preferences), within the groups of judges with different backgrounds.	189
Table 30 Standardized Deviates for the number of occurrences of ERs with positive statements to describe posters chosen during the fourth stage of the interview (personal preferences), within the groups of judges with different levels of professional experience.	190

INTRODUCTION

1 Studying creativity assessment in design

Recent years have seen a growing demand for innovation within our society. Organizations such as the European Union and the Organization for Economic Cooperation and Development emphasize the importance of creativity to the development of knowledge-based economies in order to respond to global issues like climate change, health care, economic competitiveness, social cohesion or individual well-being (Andiliou & Murphy, 2010). The European Parliament declared as a goal of the European community “to promote creativity, through lifelong learning, as a driver for innovation and as a key factor for the development of personal occupational, entrepreneurial and social competences and the well-being of all individuals in society” (European Parliament and Council, 2008, §9).

Hence, it has become vital for companies to be able to identify the extent and nature of the creative components of new products. Top innovative companies can generate over 75% of their revenue from innovative products (the ones that did not exist five years ago) (Milton, 2003). At the same time, the pace of technological change continues to accelerate, and there is fierce competition between brand names offering similar products. It is therefore particularly important for the innovative products and services that companies are encouraged to develop to have high visibility and a good external image - where *good* means creative, special, and suited to users’ changing needs. In this respect, it is crucial – not just for companies, but also for professionals who sell their creative potential - to know how their products are judged by different market players. There is therefore a strong need for measures that can distinguish between creative products and less creative ones.

The literature on creativity identifies numerous ways of measuring and fostering it. Some authors hold that measuring the creativity of the outcomes of creative work is the best way of assessing a person’s creative potential. They consider that assessing creative *products* is simpler and more efficient than assessing the creative *process*, which is long, implicit, difficult to measure and does not give any guarantee of a creative outcome.

Amabile's (1982) consensual assessment technique (CAT) has gradually come to the fore over the past few years. However, while it is widely used in a variety of creative fields, this technique seems less well suited to the particular features of design, as design tasks involve a wide range of constraints, leading to the application of more specific criteria in the assessment of design outcomes. Furthermore, the CAT's purpose is to achieve agreement between expert judges, not to explore the diversity of opinions expressed by judges with different viewpoints, which can help designers anticipate how their work will be received and provide companies with detailed judgments, rather than "all or nothing" verdicts

Several studies have highlighted differences between judgments of design products, especially between those expressed by judges with different levels of expertise in the field of design (Kaufman, Baer, & Cole, 2009; Lee, Lee, & Youn, 2005; Plucker, Holden, & Neustadter, 2008). Most of these studies, however, were seeking to demonstrate the presence (or absence) of inter-judge agreement, and the nature of the differences between these judgments, as well as the reasons behind them, were seldom analysed.

2 The field of graphic design

The world of professional design makes a clear distinction between specialist fields in design: besides the graphic design, we have product design, architectural design, fashion design and others, insofar as most of these specializations have their own occupational statuses and their own separate training courses. That said, the distinction can sometimes become blurred by the fact that many designers offer services in more than one specialist area.

The research described in this thesis covered one of the many areas of design, namely graphic design. In order to situate it within the professional context, we propose the definitions of graphic design created by a governmental institution. La Maison des Artistes (the French administrative institution for professionals in art and graphic art) indexes the activity of graphic designers, defined as aiming to transmit a visual message, within economic, social and cultural fields, using all kind of media (press, advertising, publishing, audiovisual, multimedia, ...) and all kind of technologies, with or without the use of computer tools (La Maison des Artistes, s.d.). The U.S. Department of Labor (governmental department responsible for occupational safety) describes the graphic designer as *creating visual concepts, by hand or using computer software, to communicate ideas that inspire, inform, or captivate consumers* (Bureau of Labor Statistics, s.d.).

Graphic design's means of expression, compared to product design, are reduced to two-dimensional supports. Its function consists mainly in transmitting information and building brand image, while the three-dimensional outcomes of product design are especially use- and manipulation-centred. In view of the above trends, we set out to deepen existing knowledge about this area of design, with a special focus on the way in which its creativity is perceived. Therefore, we examine the similarities and differences existing between graphic design and product design (the area of design that received much more attention within the past creativity studies), especially by identifying the key assessment criteria used by people in these two areas. Moreover, the designs used in our experiments were produced by students in graphic communication, and assessed by judges having some professional experience in this area.

3 Aims of the research

The main goal of this thesis was to analyse the factors responsible for differences in the way creative designs are assessed. More specifically, we analysed differences in judges' professional viewpoints and in the assessment criteria they took into account. Our aim was therefore twofold:

- First, we wished to identify the criteria that are spontaneously used by judges to assess design creativity, especially within the area of graphic design, by comparing them to the criteria used in the product design area. Furthermore, we attempted to find out whether there is a difference between the criteria that judges say they apply when assessing design products (their mental representations of design creativity), those they actually apply in a scores attribution task (typically used in creativity research) and those they apply in real-life product assessment.
- Second, we wished to find out whether there is a difference between the judgments made by judges with different profiles. We used two factors that could differentiate judges: their backgrounds (different professional statuses that might influence the perception of design) and their level of experience (number of years spent on professional activity within the design area). To understand the influence of the background factor, we compared the viewpoints of professional designers (professional creators of design), design teachers (professionally involved in design education), design users or audience (not involved in the creative design process, but representing its final target), and art directors (who professionally deal with the design audience preferences). To understand the influence of level of experience, we divided our sample into three categories: experts (having more than 10 years of experience within the field), intermediary-experts (having between 5 and 10 years of experience within the field) and laypeople (with no experience within the field).

4 Combining multiple methods

With these two aims in mind, we adopted an approach that encompassed several different methods. It is important to realize that creativity assessments depend on many different factors: the judge's implicit understanding of creativity; the goal of the assessment (a genuine design commission, a design contest, a personal purchase, etc.); the amount of time the judge has to arrive at a verdict; personal or professional preferences; the presence or absence of other objects with which to make a comparison, etc. In addition, collecting judgments of creativity involves asking questions, and the way these questions are framed can affect the way the judges respond. For example, judges may give one answer when asked to rate a product on a scale, and quite another when invited to express their spontaneous observations about that product. Similarly, judges may name one set of key criteria for assessing creativity, and use quite a different set of criteria in real-life judgments.

In order to gain a better understanding of all these factors, we decided to diversify our methods for analyzing creativity assessments.

First, we investigated judges' mental representations on the creativity of design, as well as the criteria they deemed relevant for assessing the creativity of design products. Moreover, we wished to know if their mental representations on creativity in graphic design are different compared to those in other design areas. To this end, we began by administering an online questionnaire in which we asked judges with different profiles to list the criteria they take into account when assessing the products in two areas of design, namely product

design and graphic design. We then conducted three successive analyses: (1) we compared the sets of assessment criteria listed for each design area; (2) for each area, we then identified those criteria that had been cited by more than 50% (and in the second time by more than 33%) of respondents; and finally (3) we compared these criteria in relation to the judges' different profiles.

Second, we looked at the criteria used in real-life design assessments in the context of the different judges' profiles. In order to obtain quantitative data, we asked our panel of judges to rate different designs on a 7-point scale. We carried out three types of analysis: (1) we identified the criteria that were most closely correlated with the overall assessments of designs (i.e., the main criteria on which the judgments were based); (2) for each assessment, we measured the levels of inter-judge agreement for each criterion and for each judge profile; (3) we compared the assessments made by judges with different profiles to see if there were any differences in the criteria they used to form their judgments.

Third and last, we collected qualitative data, in order to identify the sources of variations in the design assessments. We proceeded by collecting the verbalizations produced by the judges when they were asked about the creativity of the different designs. We analysed (1) variations in judgments depending on the judges' profiles, (2) the criteria expressed during their verbalizations, in particular, their nature and their frequency of occurrence, (3) the different points at which the criteria were used during the assessment process: first impression, in-depth analysis, final choice and expression of personal preference.

5 Thesis structure

The present thesis is divided into three parts and seven chapters. The first part presents the state of art and comprises two chapters: the first addresses the general theoretical background, the second takes a closer look at the specific context of the study.

The second part includes the third chapter about considerations related to the methodology of the research and its main hypotheses, the following three chapters present research findings. The goal of this thesis is to explore creativity criteria in different assessment contexts, thus the first study focusses on declared criteria and tries to find the common mental representations of creativity in graphic design. The second one explores the way in which judges attribute scores using previously defined assessment criteria. The focus is put on the correlations between these criteria and the inter-judge agreement on them. The third study includes the qualitative analyses of verbalizations expressed by judges during the real situation of design assessment.

The final part contains the seventh chapter consists of a general discussion and the contributions of the thesis at the empirical, theoretical, and methodological level and future perspectives.

PART I – STATE OF THE ART

Chapter 1 – Concepts of *creativity* and *design*

Our aim was to study assessments of creative designs. However, before tackling the process of assessing creativity in design, we first need to define the concepts of creativity and design, in order to understand the issues and difficulties arising from the assessment of creative design products.

In this chapter, we describe the framework within which we conducted our study, explaining the main concepts contained in the literature. We provide an overview of existing approaches to creativity, setting out the problem of how to define creativity, describing different approaches and taxonomies, and concluding with a presentation of the models that locate creativity in the social context and underline the importance of external viewpoints on creativity.

This is followed by a short state of art with regard to design and its main concepts, focusing on cognitive approaches to the design process and its various stages.

1 Creativity

1.1 Definition of creativity

Anyone working on creativity first has to define his or her understanding of this term (Plucker, Beghetto, & Dow, 2004). However, such is the complexity and multidimensionality of the subject that a clear definition is difficult to achieve.

Plucker, Beghetto, and Dow (2004) selected 90 different articles from peer-reviewed journals on creativity, business, education, and psychology, restricting their choice to those with the word “creativity” in their title. Only 38% of these articles explicitly defined creativity.

Some authors opine that creativity escapes definition (Amabile, 1996; Piffer, 2012). For example, Amabile (1996) asserts that the current state of science does not provide a sufficiently clear description of creativity for it to be given a definition. She claims that there is a plethora of data, but a dearth of definitive statements: we cannot yet establish the cognitive profile of a creative individual, that is, a person with all the traits and abilities needed to ensure the production of a truly creative outcome. Nor can we list the features that set a truly creative outcome apart from a noncreative one. However, quoting Kosslyn's (1980) observation that "it is not necessary to begin with a crisp definition of an entity in order to study it" (p.469), Amabile suggests that as long as the entity under consideration can be recognized with a reasonably good consensus, it makes sense to proceed with a scientific examination of that entity.

Many authors writing about creativity nonetheless attempt to provide some kind of definition of the term. Sarkar and Chakrabarti (2008) analysed over 160 definitions of creativity and arrived at the following *common* definition: "Creativity occurs through a process by which an agent uses its ability to generate ideas, solutions or products that are novel and valuable". This is the definition we adopted for the purposes of the present research.

1.2 Perceptions of creativity

Creativity is a subject that arouses a great deal of interest and curiosity. People are fascinated by the way in which famous creative personalities such as Albert Einstein and Pablo Picasso came up with their discoveries or produced their masterpieces (Gardner, 1993). But creativity is not only about the great art, it also concerns the everyday life, since most of the artifacts present in our environment are the consequence of people's creative invention (Bonnardel, 2006; Lubart, Mouchiroud, Tordjman, & Zenasni, 2003).

Back in the 1950s, psychologists shifted their attention from the works of geniuses to those of *ordinary* people, and the late 1980s witnessed a growing interest in the social and cultural dynamics of creativity, including in everyday life (Craft, 2005; Lubart, 1999). Increasing use is now being made of consensual forms of validation (Amabile, 1996; Hennessey, 2003), and creativity has started to be perceived of as something that takes place in the context of the community, within networks of social relations and social interactions, and using existing cultural artifacts.

The growing interest for the subject became a large public center of interest: self-help books, courses and workshops on how to develop one's creativity are extremely popular in today's society. There is an increasing demand for innovation in our society, in the form of new products, but these new products must be tailored to users' actual needs and cognitive abilities.

From the theoretical point of view, the societal approach described above contrasts personal creativity, or creativity with a small *c*, with "historical creativity" or Creativity with a big *C* (Boden, 1990; Gardner, 1993). *Personal creativity* refers to creations that of lesser importance for humanity as a whole, but which nevertheless have a great deal of value for their authors, as they are the result of an individual process. *Historical creativity* refers to the discoveries and masterpieces of famous creative personalities.

Beghetto and Kaufman (2009) extended this discontinuous view of the creativity concept by introducing the *Four C* model of creativity, adding to the existing creativities (with a small *c* and big *C*), creativity with a mini *c*, which refers to the "novel and personally meaningful interpretation of experiences, actions, and events" (Beghetto & Kaufman,

2007) and creativity with a *pro c*, referring to the creativity expressed through people's creative professions, without any major impact on history.

In lines with these hypotheses, Johnson and Carruthers (2006) divided creativity into four categories: (1) *Creation-common-place*, of which the result is neither surprising nor uncommon, just a consequence of human activity; (2) *Creation-creative-domain*, where the creative domain can be art, the media, and so on, but where the creative discovery is made on an individual scale; (3) *Creative combination*, which involves the improvement of an already existing artifact, by modifying a single feature such as method, context or use; and (4) *Creative-new*, which is innovative in the context of the history of humanity.

This desire to categorise creativity reflects the existence of different expectations associated with creativity, described by Sarkar and Chakrabarti (2008). The novel and valuable character of a product is perceived differently according to whether it has been created by a kindergarten pupil, a nonexpert adult or a creative professional.

In this thesis, in relation to the typology set out above, we focused on a brand of creativity that lies midway between the small-*c* and pro-*c* categories, insofar as the authors of the creative outcomes we analysed were students enrolled on creative courses.

1.3 Taxonomy of creativity: process, person, place and product

Creativity can be approached on the basis of Rhodes' 4Ps taxonomy: the creative *process*, the creative *person*, the creative *place* and the creative *product* (Rhodes, 1961). Hasirci and Dermirkan (2003) observed that these components, especially process, person and product, differ significantly from each other and are mostly studied separately. However, it is important to study these components together, in order to recognise the many ways in which they interact.

The creative process. Many studies have focused on the processes involved in the generation of creative solutions. The concept of divergent and convergent thinking is mainly used to explain the creative process, but there are also more specific concepts, such as the exploration of existing *conceptual spaces* (Boden, 1990), and *eureka moments* (Schneiderman, 2000).

In the literature, several models have been put forward to describe the creative process. Some of them suggest the existence of several consecutive stages (Amabile, 1996; Gelb, 1996; Wallas, 1926), and we compare these with the stages of the design process later in this chapter. The Geneplore model proposes just two main processing phases in creative thought: a generative phase and an exploratory phase (Finke, Ward & Smith, 1992; Smith, Jonides, Koeppel, Schumacher & Minoshima, 1999; Ward, Smith & Finke 1999). The generative phase involves the construction of mental representations, or so-called *preinventive structures*. These structures are then expanded and interpreted in meaningful ways during the exploratory phase. These two stages can be repeated for as long as it takes to arrive at a creative product or idea.

The creative process is often treated as a process of problem resolution. In cognitive psychology, the term *problem* refers to an individual's mental representation of his or her task (Stacey & Eckert, 2003, p.179). Mayer (1989) made a distinction between two types of problem: *routine problems* that can be solved through the application of well-known procedures, and *nonroutine problems*, where the task representation does not evoke any particular procedure for the individual, and he or she therefore has to construct a new one.

Leplat and Pailhous (1981) stressed the fact that the same task can constitute a nonroutine problem for one person and a routine problem for another person.

Consequently, in contrast to procedural activities, which have clearly defined goals and learnable strategies, creative processes involve either nonroutine problems (also referred to in the literature as *ill-structured* or *wicked* problems), which lack a clear, complete initial representation (Eastman, 1969; Reitman, 1964; Simon, 1973, 1995), or *open-ended* ones with a large number of possible solutions, all of them satisfying criteria or constraints to varying degrees (Fustier, 1989).

The creative person. Studying creativity as a personal attribute involves investigating individual differences in people's creativity and the specific characteristics of creative people (Mayer, 1999). Numerous methodologies have been developed to this end, including biographical inventories (e.g., Schaefer, 1969a; Taylor, 1975), personality inventories (e.g. Torrance & Khatena, 1970), and behavioural tests (e.g., Guilford, 1967; Torrance, 1962).

These studies may focus on different issues, including personality, intelligence or culture (Feist, 1998; Policastro & Gardner, 1999).

The creative place. Authors adopting this approach focus on the context or environment in which the creative activity takes place. Thus, teamwork, networking between employees, discussions, the external environment and task control are all studied (Amabile, 1996; Williams & Yang, 1999) as are the tools and/or media used to perform creative tasks (Do, Gross, Neiman & Zimring, 2000; Goldschmidt, 2001).

The creative product (or solution). This approach defines creativity in terms of the observed quality of the outcome of the creative process engaged in by a creative individual. By *creative product*, we mean an object whose characteristics present a certain novelty and which is adapted to a set of constraints and to the context in which it occurs (Bonnardel, 2009; Bonnardel & Zenasni, 2010; Sternberg & Lubart, 1999). According to Runco (2004), the fact that a person has a high creative potential or works in an environment that is conducive to creativity is no guarantee of a creative outcome.

In this thesis, we focus on the creative product, and this topic is therefore developed further in later sections.

1.4 System-oriented models of creativity

Sternberg and Lubart proposed a model featuring various factors that have an impact on creativity from a more global perspective (Lubart et al., 2003; Sternberg & Lubart, 1995). This multicomponent model (Fig. 1) reflects the full complexity of the influences to which creative people are exposed in the course of their work. It draws on previous research in this area and provides a detailed overview of cognitive (abilities, knowledge, intelligence, etc.), conative (personality, motivation, etc.), emotional and environmental factors. These factors interact with each other, and on many occasions the strong presence of one compensates for the lack of another. Therefore, creativity cannot be associated with any one factor, or even a specific set of factors. Creative production is the fruit of a specific and individual combination of factors that influence the creative potential of a given individual. This unique and personal aspect of creative production is reflected in the reception given to it by society.

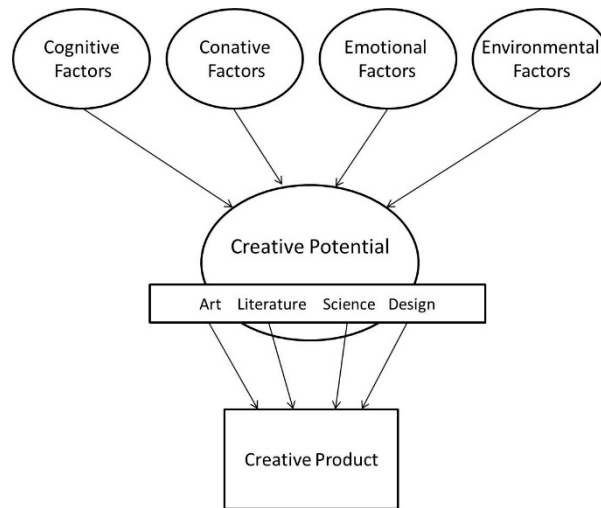


Figure 1: Multicomponent model (Lubart et al., 2003; Sternberg & Lubart 1995)

The systemic model (Fig. 2) developed by Csíkszentmihályi (1999) offers another interesting approach, whereby creativity is constructed through an interaction between the individual and a social system in which individual productions are judged by a *field* and either rejected or accepted into a *domain*. There are three associated systems: (1) the *individual*, who brings transformations into a given domain; (2) the *domain*, which consists of cultural knowledge in the form of ideas and productions selected by a field; and (3) the *field*, made up of a group of persons or institutions (*gatekeepers*) who control a particular domain by assessing and selecting those ideas and productions that should be included in it.

Thus, creativity is determined not only by the originality of an individual's production, but also by the degree to which this production is accepted and deemed to be innovative by the field. This model suggests that, in order to succeed with a creative production, an author must be familiar with the judges' requirements and values, in order to satisfy them. Thus, Simonton (1995) suggests that, in addition to Rhodes' 4Ps, we should consider the ability of creative individuals to persuade others of the value of the products they have created.

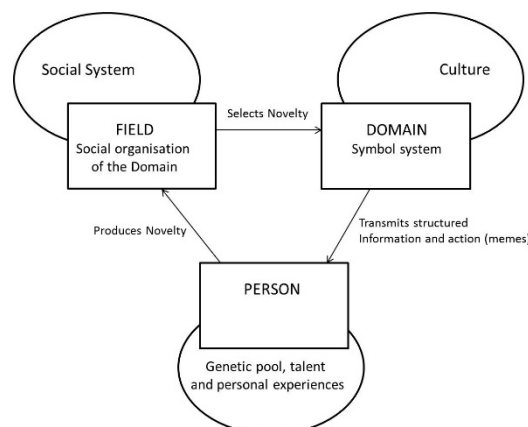


Figure 2: Systemic model (Csíkszentmihályi, 1999)

In the following sections of this thesis, we focus in detail on the process by which a creative product is accepted or rejected by the field.

2 Design

In the field of design, creativity has to meet a number of requirements and constraints. It plays a major role in design tasks, because designers use their creative skills to frame a design problem from unconventional viewpoints, explore new ideas and develop innovative design solutions (Cross, 1997; Gero, 2000; Hsiao & Chou, 2004; Van der Lugt, 2000). Designed artifacts cover a vast range of products, from everyday objects to huge buildings.

2.1 Emergence of the cognitive approach to design

Researchers have been interested in design ever since Simon (1969) claimed that design should be recognized as a cognitive activity rather than solely an occupation.

Several models and theories of design have been put forward, many of them quite close to the creativity theories. Of the two most well-known approaches to design, the first - Simon's symbolic information processing (SIP) theory - consists in considering design as problem solving. The second, represented mainly by Schön, situated activity or situativity (SIT) theory, regards design as a reflective practice or some other form of situated activity.

Symbolic information processing (SIP). According to the study by Simon and Newell (Newell & Simon, 1972) design is a type of problem-solving task, in which the *initial state* has to be transformed by *operators* (actions applied to solve a problem) in order to reach the *goal state*. In the initial state, as not enough information is provided by the environment, the designer has to define the problem space based on his or her perceptions and knowledge. Although the way in which the problem is defined affects the designer's choices for resolving it, the designer collects and generates a considerable amount of additional information during the problem-solving process, in order to find an appropriate solution. It is therefore possible to modify the initial problem space.

This approach, or variations of it, has been adopted by many authors over the years (Akin, 1986; Eastman, 1969; Goel, 1995). Nevertheless, it has often been criticized, especially for considering that the use of general problem-solving mechanisms can be directly transposed to design, and for underestimating the importance of *nondeterministic* leaps depending, for example, on analogical thinking (Visser, 2006).

Situated activity (SIT). For Schön, designing should not be regarded solely as a problem-solving activity. In his conception of design, he refers to "reflection in action" and "reflective conversation with the materials of a design situation" (Schön, 1988). Designers observe their sketches, which allow them to express or *externalise* their ideas, and which support visual reasoning. These sketches can be easily transformed, and the *unintended consequences* of these transformations may call for fresh reflection-in-action and fresh transformations. Schön talks about "improvisation on the spot" and cycles of "seeing-moving-seeing" (Schön & Wiggins, 1992).

According to this approach, the design world is constructed by the designer. The designer establishes the dimensions of the problem space and the moves needed to achieve the solution.

Dorst (Dorst, 1997; Dorst & Dijkhuis, 1995) consider that both perspectives can be useful for describing and studying design, but for different reasons. Simon's perspective on design involves more objective interpretations, while Schön's perspective involves subjective impressions and is more useful for studying conceptual design. For her part, Visser (2006) concludes that the SIP paradigm focuses on designers' use of knowledge and representations, while the SIT paradigm focuses more on the impact of environmental and sociocultural factors on the design process.

2.2 Specificity of the design process

Here we describe several models that describe the design process in terms of its specifically iterative nature and its associated activities and elements. According to Bouchard (Bouchard & Aoussat, 2003), in design process the problem space (or the design brief) is progressively transformed into the solution space (the final production). It comprises series of iterations, illustrated by the *spiral metaphor* proposed by Zeisel (1981) demonstrating the course of problem solving in design. According to this author, the design process includes numerous cycles, with the spiral becoming smaller and smaller as the designer approaches an acceptable solution.

Design tasks requires both creativity and the satisfaction of constraints. Observations of real-world design situations show that new ideas are often inspired by old situations, which may either belong the same conceptual domain, or else come from a completely different one. In Bonnardel's (2000, 2006) *Analogy and Constraint Management* (A-CM) model, two main cognitive processes interact within the design activity, frequently resulting in opposite effects.

Analogy making allows designers to draw a connection between two different concepts and to perceive their common aspects. This process can lead designers to extend (or restrict, depending on the context) their search space for new ideas (Bonnardel, 2000). The process of evoking new ideas can lead to problem resolution, as well as to changes in mental representations (Bonnardel, Didierjean, & Marmèche, 2003).

Constraint management guides designers' problem solving and helps them to restrict their search space and, by so doing, limit the number of possible design responses. As such, constraints help designers to orient their analogy making, by taking their knowledge and the design task context into account (Bonnardel, 1989, 1999).

In the literature, we find several different types of constraints: (1) *prescribed constraints*, which result from the initial formulation of the problem and can include both explicit specifications and designers' implicit interpretations (Lebahar, 1983); (2) *constructed constraints*, resulting from the designers' previous experiences and knowledge stored in memory (Bonnardel, 1999; 2012); and (3) *deduced constraints*, generated by iterative problem reformulation and the continuous evaluation of the ongoing resolution of the design problem. Additional constraints emerge to specify the properties that upcoming solutions should have (Bonnardel, 1993; 2012).

In his *Function-Behaviour-Structure* (F-B-S) model, Gero (Gero, 1990; Gero & Kannengiesser, 2004) described the different aspects that designers have to take into account in object design. This approach defines what the object is for (Function), what it does (Behaviour), and what it is, in other words, what its components are and how they fit together (Structure).

2.3 Stages of the design process

Several authors conceive of creativity and the design process as a series of stages. Creativity occurs mainly during the initial stages of the design process, in which we can observe the appearance of creative ideas (Bonnardel, 1999). If we compare the design process (Fig. 3: Asimov, 1962; McNeill, Gero, & Warren, 1998) and the stages of the creative process (Fig. 4: Amabile, 1996; Gelb, 1996; Wallas, 1926), we find a great many similarities. Thus, we can identify three main activities in common: problem formulation (and reformulation), solution generation and solution evaluation (Visser, 2006).

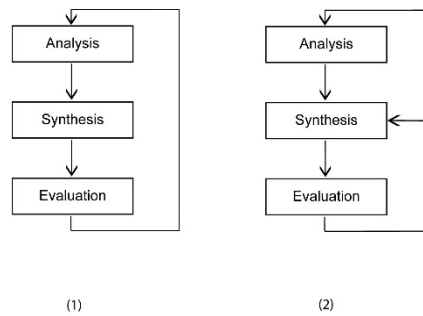


Figure 3: Design stages: (1) Asimov, 1962; (2) McNeil, Gero, & Warren, 1998 (from Bonnardel, 2009)

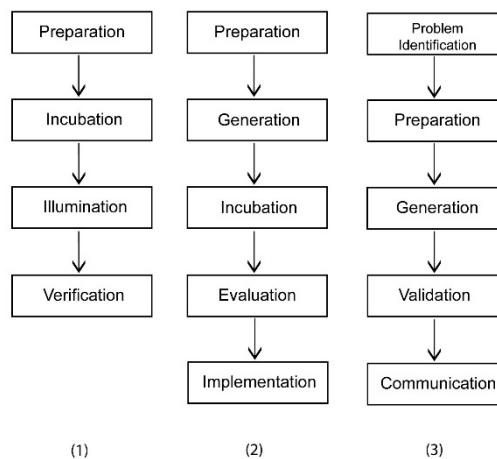


Figure 4: Creative stages: (1) Wallas, 1927; (2) Gelb, 1996; (3) Amabile, 1996 (from Bonnardel, 2009)

These activities are neither linear nor independent. Visser concluded that the formulation and generation processes often take place in parallel (Visser, 2009a), and that generation is often closely correlated with evaluation. Several authors mention the opportunistic organisation of design activities, in which each decision is motivated by the one before (Bonnardel, Lanzone, & Sumner, 2003; Hayes-Roth & Hayes-Roth 1979). As a consequence, the decision process in design can be multidirectional, resulting both from top-down and from bottom-up processes. It can also take place at different levels of abstraction.

2.3.1 Formulation and reformulation of the problem

Most of the above-mentioned models propose *preparation* or *analysis* as the first stage of the design or creative process. This includes the search for relevant information and the formulation and reformulation of the design problem.

Numerous studies have shown that, owing to the complexity of the design problem and the various task constraints that have to be taken into account, designers spend a considerable proportion of their time refining their mental representation of the design problem. Often, they focus more on problem formulation than on problem solving (Bonnardel & Sumner, 1996). This problem representation depends a great deal on the designer's prior knowledge and expertise (Bonnardel, 2009).

As we mentioned earlier in our description of the creative process, one of the main characteristics of design problems is that their initial state is ill structured (Eastman, 1969; Reitman, 1964, Simon, 1973, 1995). The designer's mental representation becomes clearer as problem solving progresses and the search space for potential solutions is gradually restricted until the designer arrives at a solution that he or she believes satisfies certain criteria. In the literature, this process is described as the *co-evolution* of problem and solution spaces (Dorst & Cross, 2001) or as an iterative dialectic between problem framing and problem solving (Rittel & Webber, 1984; Simon, 1995).

2.3.2 Generation of solutions

The next stage is referred to by authors in terms of *generation* or *synthesis*. It can include specific substages, such as *incubation*, when different ideas unconsciously emerge, and *illumination*, when these ideas are consciously discovered (Gelb, 1996; Wallas, 1926).

A solution seldom appears out of nowhere. Designers generally generate solutions from their generic knowledge or engage in analogical reasoning, returning to similar problems they have successfully solved in the past (Visser, 2002, 2010). Generation may involve transformation, which, depending on the degree of proximity between the source and the new representation, can consist in duplicating, adding, detailing, concretizing, modifying or revolutionizing (Visser, 2010). This is why some outstanding designers can be classified as *solution-* rather than *problem-based* in their search for ideas (Cross, 2004). They have more examples stored in their memory that they can reuse and develop in new design tasks than novice designers.

To generate solutions, designers perform a synthesis of the analogies and constraints they took into account in the problem-formulation phase. The previously defined aspects and properties become the criteria that guide them in their search for solutions. In the resolution of the design problem, constraints restrict the search space for solutions, whereas criteria give directive principles (Bonnardel, 2006). This means that designers can work at different levels of abstraction to generate solutions, according to the advancement of their work (Darses, & Détienne, 2004).

The externalisation of ideas and their assessment are interdependent processes in design (Bonnardel, 2006). *Externalisation* refers to the creation or modification of external representations (sketches are the most common examples of this process) and is highly influenced by context (Wojtczuk & Bonnardel, 2010). It allows designers to engage in a dialogue, or *reflective conversation* with their project, which serves as a basis for reasoning and exploring, as well as for criticizing their ideas (Schön, 1983). Furthermore, it allows them to communicate these ideas to other people. According to Schön's (1983) concept of *reflection in action*, designers may make unexpected discoveries about their work while

externalising. These discoveries can be positive (results perfectly meeting requirements), negative (emergence of problems interfering with the goals being pursued) or innovative (perception of new directions for creative research). To allow for the advent of these unexpected elements, representations produced during the early stages of design need to be abstract, ambiguous and imprecise (Gross & Do, 1996; Tversky, 2003). As long as these representations remain abstract, the specification of details can be left until later.

2.3.3 Evaluation of solutions

The process of idea generation is associated with the stage known variously as *evaluation* (Gelb, 1996), *verification* (Wallas, 1926) or *validation*, followed by *communication* (Amabile, 1996). To evaluate their ongoing work, designers use evaluative referents, based on the constraints defined during the problem-formulation phase, as well as any criteria that have arisen from external priorities or internal reflection and preferences (Bonnardel, 1996; 2000; 2012; Visser, 2009a).

The evaluation process allows the designer to select solutions that satisfy specific constraints and to exclude the inconvenient ones. To do this, the designer assigns a different status to each of the evaluative referents and defines their relative importance for the designed object (Bonnardel, 1999). The designer can judge the idea positively or negatively, or validate the direction taken for the work in progress. Most design decisions are based on some form of evaluation, as described in the theory of planning as an opportunistic process (Bonnardel, Lanzone, & Sumner, 2003; Hayes-Roth & Hayes-Roth, 1979).

The major problems encountered in the course of design evaluation result from (1) *incomplete evaluative knowledge*, meaning that designers with little professional experience lack knowledge and use fewer criteria (Bonnardel, 1991), (2) the *use of multiple perspectives*, or different points of view, implying the existence of specific goals, relevant bodies of design knowledge and preferred forms of solutions, and (3) the *recognition of problematic solutions*, when a flaw is discovered in the current design solution and has to be overcome (Bonnardel, 2006).

Three types of evaluation are possible (Bonnardel, 2006, 1993): (1) *analytic*, whereby the positive and negative sides of solutions are identified, using only the evaluative referents; (2) *comparative*, whereby alternative solutions or specific solutions are measured against a benchmark; and (3) *analogical*, whereby previous evaluations of similar solutions are transferred to the one under consideration.

The evaluation process described above basically consists of self-evaluation performed by the designers themselves. It is important for designers to assess their work, as it helps them to choose the best ideas for continuing it (see Bonnardel, 1999). However, several studies have shown that self-evaluation is not always compatible with the evaluations made by others (Kaufman, 2006; Priest, 2006), and the following chapter tackles the subject of evaluation other than from the author's viewpoint.

Chapter 2 Creativity assessment

The second chapter of this thesis focuses on assessment of creative design by external judges. In the following chapters, we use terms “judgment” and “assessment” as synonyms, to designate the activity of collecting relevant information that may be relied on for making decisions (Fenton, 1996). In our thesis, making a decision consists in attributing a score or expressing an opinion and is performed by judges. Accordingly, we use the term “judge”, to designate every person who may be in the situation of expressing an assessment.

We will not deal with the situations, in which the assessment concerns the judges’ own solutions (this was described in the previous chapter of this thesis and concerns the evaluation as a stage of design process). We focus on the situations, in which a judge assesses a solutions produced by somebody else. The examples of this second situation are multiple: we can cite the educational context, in which a teacher assesses students’ work; the professional context, in which an art director assesses his or her team-members’ productions; the commercial context, in which a potential customer assesses the products in order to decide if he or she is interested in purchasing them or not.

In the following chapter, we focus on the assessment of creative products. After providing an overview of creativity assessment methods, we take a closer look at the assessment of creative products. We have chosen to focus on a particularly popular approach to creative product assessment, based on the opinions of external judges: the CAT. We discuss the technique’s strengths and weaknesses in the context of design.

The second part of this chapter, is based on the Brunswik lens model of judgment (Brunswik, 1955), which, put in connection with creativity assessment, invites us to tackle the two factors that we believe merit in-depth analysis with regard to design creativity judgments. First, we focus on the different viewpoints that judges can have towards design products, taking their experience within the field of design into account. Second, we provide an overview of the criteria used to assess the creativity of design products in the literature.

Many researchers studying creativity aim to establish creativity measures. We present here a quick historical overview of this research field and the main creativity measures that are actually used within creativity studies.

1 Taxonomy of creativity measures

Interest in assessing creativity has increased in recent years. Assessment is an integral part of education and, indeed, of any endeavour to improve creativity in one’s professional or personal life. In the following section, we recall some of the most important points in the development of creativity measures. As mentioned in the previous chapter, creativity is difficult to define and, as a consequence, seems difficult to measure and to assess. Despite this, there is a long tradition of measuring creativity.

The taxonomy of creativity measures can be likened to the taxonomy of creativity itself (discussed in the previous chapter). Thus, we can distinguish between four groups of creativity measures (Plucker & Makel, 2010): (1) *creative process*; (2) *creative person*; (3) *creative products*; and (4) *attributes of creativity-fostering environment*. After briefly describing these measures, we turn our attention to the main subject of this thesis, namely product creativity assessment.

Creative process measures mostly comprise divergent thinking batteries. The idea behind these creativity tests is to ask participants to produce as many solutions to a problem as possible, and to measure their fluency (by counting the number of produced ideas), flexibility (by taking into account the number of ideas categories), originality (by checking their frequency of occurrence within the same population) and elaboration (by measuring the amount of detail in the responses). These tests can be applied to all possible fields, like graphic creativity, verbal creativity (Wallach & Kogan, 1965), social creativity (Mouchiroud & Bernoussi, 2008) etc. The best known batteries are Guilford's (1967) Structure of Intellect and Torrance's (1962, 1974) Tests of Creative Thinking. Divergent thinking tests have the weakness of measuring only a part of the creative process characteristics. If we look at the multicomponent model by Sternberg and Lubart (1995), presented in the Chapter 1 of this thesis, we can see that there are many other creativity components (the cognitive, conative, emotional and environmental ones) that are not taken into account by divergent thinking tests.

Creative Person measures include studying individuals who have already been identified as creative, and determining their common characteristics. This group of measures comprises personality scales such as What Kind of Person Are You? (Torrance & Khatena, 1970), activity checklists such as the Alpha Biological Inventory (Taylor & Ellison, 1966, 1967), and measurements of attitudes, such as those developed by Basadur and colleagues (Basadur, Taggar, & Pringle, 1999; Runco & Basadur, 1993). The weakness of these kinds of measures is that the correlations between these measures and scores attributed to creativity are rather weak, around $r = 0.35$ (Lubart, 2005). Again, conative characteristics should be taken into account during the creativity studies, but we should combine them with the cognitive, emotional and environmental ones.

Attributes of creativity-fostering environment measures involve identifying environmental variables liable to be related to creativity, as in the KEYS: Assessing the Climate for Creativity instrument (Amabile, Conti, Coon, Lazenby, & Herron, 1996). This viewpoint on creativity, as the previous ones, is only partial. Nevertheless, it is an interesting approach in the context of facilitating creativity in the educational or professional environment.

Creative product assessments allow creativity to be defined according to the observed quality of the end products. The best known techniques of creative product assessment are the CAT (Amabile, 1996), the Creative Product Semantic Scale (Besemer, 1998; Besemer & O'Quin, 1999) and the Student Product Assessment Form (Reis & Renzulli, 1991). This form of assessment allows judging a finished product, not just the idea sketches. Therefore, it can be perceived as a final result of interactions between all the creativity factors, which makes this measure the closest to the ecologically valid creativity assessment. We will develop this subject further on in this chapter.

2 Assessing creative products

Product assessment is the main focus of this thesis. Moreover, we concentrate on product assessment in the field of design, which involves the consideration of functional issues and constraints (often clearly described in the brief), which are much less present within the assessment of pure art (based especially on aesthetical considerations).

Below, we discuss the major strengths of this method of assessing creativity as well as its different measures, with a special focus on the Amabile's Consensual Assessment Technique (1982).

2.1 Why focus on product creativity?

Here we explain the utility of focusing on product creativity measures, first from a scientific perspective, second from a practical point of view.

There are many scientific arguments to support our choice of product assessment as a means of measuring creativity. In their study, Demirkan and Hasirci (2009) showed that the product is the strongest factor (45.85%) for determining the amount of creativity involved in design. Person and process come second (19.54%) and third (14.46%). For their part, Plucker and Makel (2010) explain that “if one goal of creativity psychometrics is to predict who is most likely to produce creative works in the future, being able to create such products in the past or present would appear to be a key indicator”.

Baer, Kaufman and Gentile (2004) also claim that product assessment is the most appropriate means to assess creativity, as it avoids measuring skills that are only partially linked to creativity (often included in measures of person- and process-creativity). The focus is entirely on the actual creativity of the outcome of the creative process. In design, we can always infer a creative process from a product, whereas the observation of the design process does not guarantee that a creative *event* will occur (Dorst & Cross, 2001).

For companies involved in product development, the final outcome is the most important part of the whole creative process. Identifying the most creative version of a new product can allow a company to succeed and outstrip its rivals. Besemer and O’Quin (1999) argue that people who are involved in the process of the products’ development and evaluation, especially in the business world, need adequate, reliable, and valid means for evaluating different characteristics – including the creativity – of the products brought to market.

2.2 Existing measures of product creativity

As with every other measure, creativity measures need to be reliable (yielding similar values if repeated) and valid (measuring creativity per se rather than, for example, only aesthetics or functionality). If they are to be implemented in applied fields, they also need to be efficient (low cost in time and money) and robust (reliable and valid for a wide variety of products). Moreover, if the aim is to increase knowledge about creativity judgments, measures should reveal *why* judgments are positive or negative.

The literature on design products intended for consumers contains measures used in usability or marketing studies, rather than creativity measures. This is due to fact that the main goal of these artifacts is to meet customers’ needs. While these features are important, they cannot yield valid measures of product creativity, as they refer only to the *appropriateness* aspect of creativity, and the definition of creativity includes *originality*, as well as appropriateness. According to Christiaans (2002), there is an expectation that design products will be original and add value to existing products. This underlines the importance of measuring the originality aspect of design products, too, and consequently the utility of implementing overall creativity measures of design products.

Some authors suggest that, in order to measure creativity, we should observe the judges’ reactions, to see whether they express the *shock of recognition* (Cropley, Kaufman, & Cropley, 2011). Jackson and Messick (1965) specified that these reactions should include surprise, satisfaction, stimulation and savoring. Nevertheless, it seems rather restrictive to assess products’ creativity solely on the basis of emotional reactions, which are quite complicated to measure (efficiency problem) and can be biased by many different factors

(reliability problem). Furthermore, as explained later on in this chapter, emotional reactions are only part of a design assessment.

Most existing product assessment models propose lists of criteria that a product should meet (these criteria are described in greater detail further on in this thesis): Taylor's Creative Product Inventory (1975); The Creative Product Semantic Scale developed by Besemer and O'Quin (1987, 1999); The Student Product Assessment Form (Reis & Renzulli, 1991), intended specifically for the educational context; The Propulsion Model (Sternberg, 1999; Sternberg, Kaufman, & Pretz, 2002, 2003).

In the creativity assessment literature, we often come across the *criterion problem*, which is a direct consequence of the desire to achieve a clear definition of creativity. Yet again, we come up against the problem of validity. If we accept the basic definition of creativity as a process that allows products to be generated that are both new and valuable, we should be able to find appropriate criteria for assessing newness and usefulness. However, we also need to ensure that these criteria are relevant and understandable for judges, who need to use them with transparency and objectivity. The CAT, proposed by Amabile and described below, is an elegant solution to this problem.

2.4 Consensual Assessment Technique (CAT)

Amabile first proposed the CAT in 1982, and since then it has increasingly been used as a reliable and valid measure of creativity. It has even been described as the gold standard of creativity assessment (Baer & McKool, 2009). The CAT's validity was established empirically and it therefore does not depend on the validity of any particular theory of creativity.

2.4.1 Theoretical background to the CAT

Amabile argued that the problem of finding an objective criterion on which to assess creativity is impossible to resolve given the current state of the literature. Moreover, judgments of creativity are necessarily subjective. It was on these premises that Amabile (1982) came up with the following operational definition of creativity: "A product or response is creative to the extent that appropriate observers independently agree it is creative. Appropriate observers are those familiar with the domain in which the product was created or the response articulated." (p. 1001). By taking this as a basis for the CAT, Amabile moved away from the notion of objectivity in creativity assessment and allowed for *correlated* subjectivity.

The reliability of the CAT approach has been examined in numerous studies. Owing to its simplicity and the high levels of inter-judge agreement, it has become a popular methodology in many areas of creativity. Numerous studies have been conducted with different participants and in different areas of creativity. Below, we provide several examples of CAT use in creativity assessments with high inter-judge reliability, typically between .70 and .90:

- Ruscio, Whitney, and Amabile (1995), for assessing students' haiku poems;
- Amabile, Philips, and Collins (1994), for assessing professional artists' portfolios;
- Conti and Amabile (1995), for assessing computer programmes;
- Baer, Kaufman, and Gentile (2004), for assessing personal narratives;
- Dollinger (2007), for assessing drawings and essays;
- Plucker, Kaufman, Temple, and Qian (2009), for assessing films.

The CAT has also been used in studies exploring cross-cultural or cross-ethnic perspectives on creativity judgments (Baer, Kaufman, & Gentile, 2004; Chen, Kasof, Himself, & Greenberger, 2002). Some authors have tailored the technique to suit their research goals, by varying the judges' expertise (e.g., Dollinger & Shafran, 2005) or applying a multi-item scale (e.g., Lee, Lee, & Youn, 2005).

2.4.2 CAT limitations in the context of design

Although the CAT is extremely popular in creativity research, it should not be perceived as a readymade measure for every context or area of creativity.

For a start, the CAT may not be convenient for every type of creativity. Baghetto and Kaufman (2009) suggested that the different types of creativity defined in the Four C model (see Chapter 1) should each be evaluated with a different type of measure. Thus, self-assessment would be the most appropriate method for judging mini-*c* creativity, while major prizes and honours, or historiometric measures would be more suitable for big-*C* creativity. Consensual assessments would be best for small-*c* and pro-*c* creativity.

There also seems to be a degree of reluctance about using the CAT in the design field. A recent survey of numerous scientific journals revealed that the number of CAT citations in design journals was limited compared to other journals specializing in creativity (Jeffries, 2012), even if inter-judge reliability in design creativity studies using the CAT is above the standard 0.7 level (Christiaans & Venselaar, 2005; Pektaş, 2010). Nevertheless, CAT measures do not provide any qualitative information about the judgments that are reached. There is no risk of imposing the wrong criteria on judges, but by the same token, we have no idea which criteria guide their judgments.

Moreover, in the context of design, only taking experts' viewpoints into account can be reductive: in real-life situations, product assessments are often formulated from viewpoints other than those of experts. The same product can be evaluated by people from many different backgrounds, who probably reach different conclusions about its creativity. In any case, who should we regard as experts? Some research considers design teachers to be experts. But do they have the same approach as professionals with a more practical attitude?

In the field of design especially, criteria other than creativity may enter into product assessments. The opinions and degrees of satisfaction of the products' users, beholders or targeted public are also very important.

External judges' assessments are mostly expressed in the form of ratings. The problem is that rating scales may incorporate not one but several criteria, and these criteria mostly remain implicit for raters (Kreitler & Casakin, 2009). Creativity assessments are never one-dimensional, and rely on a complex system of beliefs (Glăveanu, 2012).

The CAT is based on agreement and consensus between judges. The final score is the average or most frequent one, and the convergence of the judges' assessments is the prerequisite for validity (Amabile, 1996; Kaufman, Baer, & Cole, 2009). The problem is that homogeneity (in terms of assessors and assessments) is not a feature of ecological, real-life situations.

In our research, we therefore decided to use a combination of methods to measure product creativity. We wished not only to collect information about the criteria underlying creativity judgments, but also to ascertain whether these criteria vary according to the

social context in which these judgments are made. In order to gain a deeper understanding of creative product judgments, we chose methodologies that would yield both quantitative and qualitative data. Consequently, in the following section, we analyze the nature and use of criteria, taking assessor heterogeneity into account.

3 The model of judgment

As we could see in the previous parts of this chapter, in order to assess the creativity of a product, most researchers rely on the experts' judgments. Our goal is to understand these judgments as well as their common and differentiating points. In order to have a better view of the factors that influence creativity assessments, we present a general model of judgment and some elements from the psychological field of decision making that we deemed relevant to our subject.

3.1 Lens model of judgment

We propose as a model of judgment, Brunswik's lens model (Brunswik, 1955), which provides a framework for investigating the nature of human judgment. It describes judgment as the set of relationships between environmental variables and how an individual perceives them to finally make a judgment. The environment (Criterion) is perceived by a person (Judgment) toward the complexity of cause-effect relationships between its elements (Cues). Thus, the cues are "the lens" through which a judge perceives the environment and is able to deduce its functioning. The goal of studying judgments is to understand the degree to which the environment and the judge's vision thereof are consistent (Achievement).

In the case of creativity judgments, our goal is to understand how a judge uses cues in order to make his or her judgment. Therefore, first (1), we need to know what are the cues used to assess creativity. For this, we proposed an overview of creativity research that uses different criteria on which creativity can be assessed (in the following parts of this chapter). We also conducted a study of criteria-finding, precisely in the field of design (Chapter 4). Second (2), we need to study the nature of the relations between these criteria and their level of importance to judges. In order to understand these issues within assessments of creativity, we should remind the reader of the characteristics of responses to the creative design problems, discussed below.

3.2 Use of heuristics

As we have seen in Chapter 1, judges have to deal with responses to open problems, in which a multitude of answers is possible, all of which can be different from each other, meeting the relevant criteria to various degrees; it is hard to judge which of these criteria are the best. Moreover, the circumstances of creativity assessment matter. In the experimental context of creativity studies, as well as in a professional context, judges typically have only a small amount of time to express their judgment. Likewise, in an experimental context judges are likely to have only a limited amount of time when giving scores to the multiple examples of design. In the professional context, some decisions must also be taken quickly and in the context of uncertainty. Indeed, as we have mentioned, there is never a single best solution to a design problem and different solutions may respond to various problem constraints to a greater or lesser degree.

In such situations, people can apply what Simon has called a concept of bounded rationality (1956). This suggests that rather than searching for the optimal solution, people are satisfied with solutions which are seen as good enough, without using all the available information and cognitive resources. Moreover, in line with Gigerenzer (2008), when judgements need to be made quickly and/or when the likelihood of different outcomes is unknown (and as we have seen, this is often the case for creativity assessments), people can apply strategies known as heuristics to their assessments. Heuristics are cognitive shortcuts that make the judgement process easier (Tversky & Kahneman, 1974). Availability is one of the most commonly described heuristics. It concerns situations, where people determine the likelihood of an occurrence based on the availability of examples that can be easily retrieved in their memory.

Thus, it is likely that, when making their assessments on design creativity, judges apply the availability heuristic and, consequently, use the creativity criteria that are the most available in their memories. Therefore, their assessments are, to some degree, shaped by their past experiences. In the following part of this chapter, we discuss which factors can be taken into account to study how judges' past experiences can influence their creativity assessments.

4 Influence of judge characteristics on creative product assessment

Csikszentmihalyi (1996) suggested that creativity depends not on the product itself, but on the way in which a person perceives the products' originality and potential to respond to his or her actual needs. This means that creativity should be perceived with regard to the context and the characteristics of its judges. Several authors (Csikszentmihalyi, 1988; k) claimed that creativity judgments should be viewed as contextual, relying on shared agreement on specific meanings within a group of people.

Norman (1988) focused on the differences between the designers' and the users' perceptions of the same product, arguing that these differences are due to the fact that their past experiences with that product were of a different nature. In line with this approach, we suggest another factor to be considered as a source of variation within the design assessments: the **judges' backgrounds**. By the term *background* we mean the judges' past experiences shaped by their professional lives and possibly influencing their perception of design.

Moreover, many authors have suggested that differences between individual judges could lead to differences in creativity ratings (Kaufman Baer, Cole, & Sexton, 2008; Lee, Lee, & Young, 2005; Plucker, Holden, & Neustadter, 2008). Typically, the judges' **level of expertise** (measured in the number of years actively spent on experimenting within the domain) is widely discussed in the literature.

4.1 Assessing design: differences in mental representations of the same product

Studies show that creativity judgments are not necessarily consensual. However, rather than regarding this as a shortcoming, we view it as an indication of an interesting phenomenon, insofar as disagreements over creativity provide an opportunity to examine the factors behind these divergent assessments.

Norman (1988) noticed that the same product can be perceived in a totally different light by its creator (the designer) and its user. His aim was not to underline the difference in expertise between professionals and novices. He focused on the background of the assessors and the consequences this had on their mental representation of the product. The influence of judges' backgrounds on the product design assessments was explored by Wojtczuk and Bonnardel (2012). They found out that the way in which specific criteria impacted the overall design assessments could be different depending on the judges' different profiles. These findings are in line with Glăveanu, (2010), who suggests that creativity, as well as its assessments, are rooted in the social and cultural context of participants. Glăveanu proposed to study the creativity assessments (within the field of traditional craft), using a multiple feedback methodology, in which the groups of judges with different backgrounds and focusing on the impact of their professional experience and position occupied in society on their judgments. Between different profiles, he differentiated the following ones:

- those involved in the creation;
- those who use the creation;
- those who are eventually interested in purchasing the creation;
- those who distribute or sell the creation;
- those who are considered as experts in the domain of the creation.

Difference between the position occupied by a person, as well as his or her background can influence the way of building the mental representations. *Mental representations* have often been studied in psychology. Richard (1995) defined them as circumstantial constructs, elaborated in a specific context and with specific goals. These representations can be influenced by numerous factors, which vary with the individual. According to Johnson-Laird (1995), people create their internal representations of reality and their so-called *mental models* based on their perception or imagination of the state of things. Similarly, Leplat (1985, 1997) talks about "fictional representations", which he describes as selective, subjective, distorted and unstable.

In line with these observations, we suggest that the differences between Norman's users and designers are a consequence of different distortions of the mental images they have constructed. Designers' interpretations are influenced by their professional creative experience, or by their knowledge of their specialist area, while users base their interpretations on their experiences of use and on the opinions either of other users or of domain gatekeepers. Nevertheless, in design, the users' or targeted public's viewpoints cannot be ignored, as it would be a huge commercial mistake to rely exclusively on the experts' judgments on the grounds that they have greater expertise. In consequence, some authors underline the importance of studying not only the models of creativity, but also of understanding the mental representations about the creativity, especially differences between the scientific concepts and laypeople's beliefs about the nature of creativity, about creative individuals or about a creativity-fostering environment (Andiliou & Murphy, 2010).

Based on the existing literature, we described below the mental representations of people involved in the design domain: designers, gatekeepers and users or audience. In lines with these characteristics we can build hypotheses concerning their criteria of creative design assessment.

4.1.1 Designers' judgments

For designers, evaluation is a part of the creative process (see Chapter 1). The main difficulties observed in such a situation are that: (1) it is hard to know all the relevant criteria, and (2) the design needs to be assessed from multiple perspectives, which is complicated by the fact that designers use their own representation of the problem to perform their creative work (Bonnardel & Sumner, 1996). To make design assessments, they take into account the interactions between project constraints and the evaluative referents and they attribute to them different levels of importance (Bonnardel, 1996). According to Lera (1981) this level of importance is based on personal values or the project's explicit specifications.

Lera's findings suggest that designers with a great deal of professional experience have a fixed view of their priorities, while novices change their priorities during the creative process. Moreover, it seems that a lack of inter-judge agreement was observed between professional designers evaluating other people's design proposals. This could be a consequence of these differently weighed attributes. From a practical point of view, designers need to be as well briefed as possible, so that they respect their customers' values and are able to agree with them on priorities. Where this is not the case, designers continue to rely on their own values and their assessment of the product may conflict with that of their customers.

4.1.2 Domain gatekeepers' judgments

Domain gatekeepers are those people, who, in the creative ecosystem, have a position that allows them to decide which creative products should enter into the domain, by being transmitted to relevant audiences (Hayes, 1989). In other words, they add memes to a domain (Csikszentmihalyi, 1999). *Meme* is a term introduced by Dawkins (1976), who gave it a role in the cultural evolution analogous to role of genes in biological evolution. Memes can convey the instructions for action (e.g. principles of aesthetics) and, in contrast to the inborn genes, are transmitted only by learning.

They are given the proof of their expertise by their own accomplishments and occupy the positions of teachers, critics, editors, museum curators, agency directors or foundation officers (Hooker, Nakamura & Csikszentmihalyi, 2003). As a consequence of their experience gained through their own creative or gatekeeping activities, they have access to a large amount of domain-relevant knowledge, which, in line with Amabile (1996), includes facts, circumstances and issues concerning a given area. Thus, domain gatekeepers should have enough domain-relevant knowledge to identify the required degree of appropriateness and of novelty.

Gatekeepers constitute the social organization of the domain. In design, their role consists in promoting creative solutions and initiating their diffusion and the evaluative process of determining what is creative can have a bottom-up or a top-down direction (Sosa & Gero, 2004). We can distinguish two groups of domain gatekeepers. The first group is composed of those who, through interacting with designers have a decisive role in adding memes into the domain. In this group we can find people occupying positions in the professional world, such as editors or art directors. The second group consists of those who actually judge the creative potential of designers and their role consists often in transmitting the already existent domain-relevant knowledge to others. We find them especially in the education world, potentially in all kind of juries.

The first group (e.g. art directors), by introducing new creations to the domain and especially to the audience, is to some degree dependent of other fields that are of importance to this audience, such as religious, political, economic or ethical considerations. The judgments that respect this dependence increases the field's effectiveness and social credibility (Csikszentmihalyi, 1999). Art directors' responsibilities are described as caring about "the overall look and feel (...), typography, visual interface design, color palette standards, page layout details and the particulars of how the graphics, photography, illustration and audiovisual elements (...) come together to form an integral whole. (...) person responsible for assuring that the new (...) design work is consistent with any established corporate identity and (...) standards" (Lynch & Horton, 2008). There is a lack of studies focusing on the way in which this group of judges perceives creativity, but we can suppose that public-appropriateness and aesthetics should be a high-priority criterion for them.

The second group (e.g. design teachers) can be perceived of as experts in assessing other people's work. Their judgment is focused on the student's learning process and individual progress (Birenbaum, 2007). Often, their assessment comprises a set of different criteria, some of them even proposed as a result of numerous studies, as in the case of the Student Product Assessment Form (Reis & Renzulli, 1991). This instrument aims, in an educational context, to assess students with regard to an early statement of purpose, problem focusing, level of resources, diversity of resources, appropriateness of resources, logic, sequence and transition, action orientation, and audience. It also includes an overall assessment containing the following sub-factors: originality of the idea; achieved objectives stated in the plan; advanced familiarity with the subject; quality beyond age or grade level; attention to detail; time, effort, energy; and original contribution.

4.1.3 Laypeople or users' judgments

The design users' viewpoints on creativity can be different from those of design professionals. This has been described in Norman's model (1988), mentioned above; moreover, several authors underline the difference between *explicit* theories held by researchers and specialists in the field, who must articulate, test and share ideas and *implicit* theories. Chan and Chan (1999) described implicit theories about creativity, as the structures built by laypeople, based on their belief systems and on self-related interpretations like personal experience, everyday knowledge or feelings (Leder, Belke, Oeberst & Augustin, 2004).

Design creativity is considered an important value within consumers' experience of a product, especially in the context of today's market, which has shifted from product-based to value-based competition (Horn & Salvendy, 2006a). It was demonstrated that creativity plays a role not only in the satisfaction of users, but also in their intention to purchase a product (Horn & Salvendy, 2006a, 2006b, 2009). Their studies on website design clearly show that creativity lies in the eyes of the beholder, which means that its assessment highly depends on the judge and the context of his or her interaction with the assessed design. Moreover, the creativity can be revealed in terms of different dimensions, depending on the context in which the interaction with the design took place.

Figure 5 presents the model of information processing during the assessment of product creativity (Horn & Salvendy, 2006b; Zeng, Salvendy & Zhang, 2009). In line with this model, users first proceed with their senses (visual, auditory or tactile ones) while interacting with the product, which will influence their affect. Then, the understanding that they have of design attributes shapes their perception of the product. After comparing the

perceived design and the context information with their criteria of creativity, the users are able to judge the creativity level of the object and decide of their response toward it.

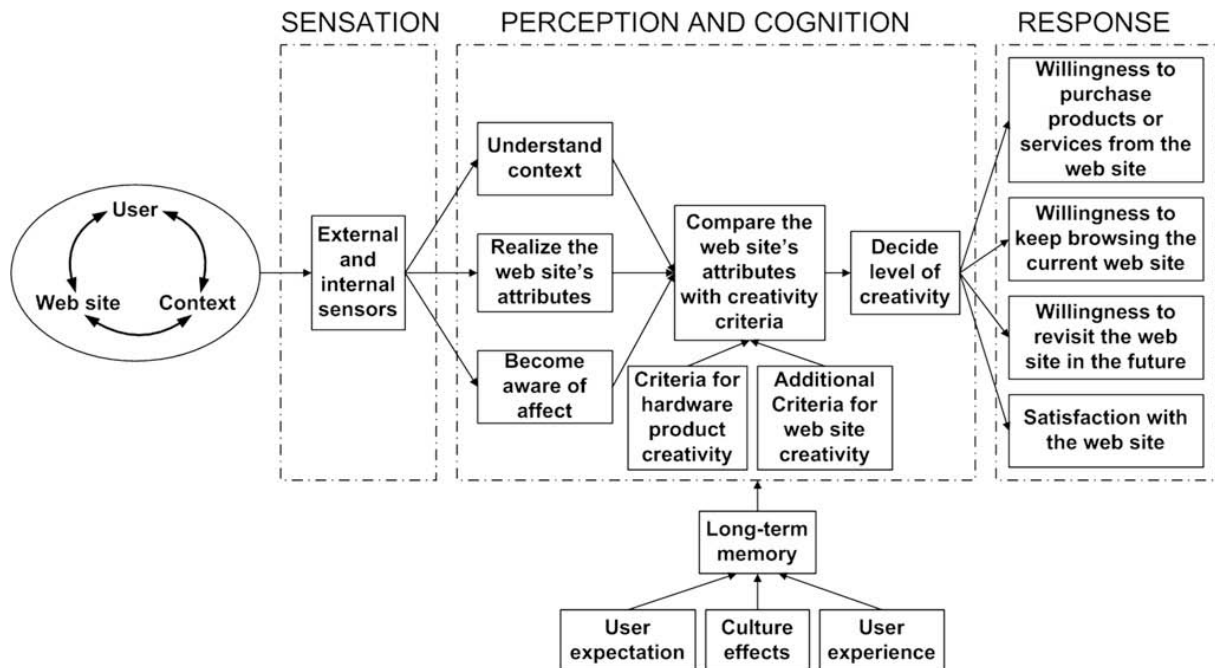


Figure 5: The information processing model (Zeng, Salvendy & Zhang, 2009).

4.2 Differences in creativity assessments depending on the judges' level of expertise

The CAT recommends choosing expert judges for assessing creative works, the argument being that assessors who possess some expertise in the relevant area will spontaneously recognize whether or not the work is creative (Amabile, 1996).

This notion of solely using experts to assess creative products has often been discussed and subjected to experimental conditions.

Authors suggest that 10 years of active experimenting and exploring the field are needed to become an expert (Bloom, 1985; Gardener, 1993). Martindale (1990) emphasizes that this expertise takes longer to acquire in creative areas, for in order to maintain the audience's interest, artists must constantly change their style to prove their creativity. Simonton (2000) not only suggests that an initial amount of time is needed to learn the mechanics of a given area of the creative field, but also postulates that a separate amount of time has to be spent on complex specialization and versatility in order to ensure creative development.

For design, Lawson and Dorst (2009) have proposed a model comprising novice, beginner, advanced beginner, competent, expert, master and visionary phases. In this model, graduate students are perceived of as advanced beginners, while visionaries are able to introduce new concepts to the profession.

Is it really necessary to use the experts with at least ten years of experience to assess creativity? Some authors (e.g., Lee, Lee, & Youn, 2005) underscore the importance of expertise in creativity assessments, especially for knowledge-rich areas. Generally speaking, inter-judge agreement is lower for nonexpert judges (Kaufman et al., 2008; Lee,

Lee, & Young, 2005; Plucker, Holden, & Neustadter, 2008). Moreover, novice ratings usually have low correlations with expert ratings (Hickey, 2001; Kaufman et al., 2008), although these correlations become stronger when studies concern more accessible areas (Kaufman, Baer, & Cole, 2009). To compensate for this lack of agreement, Dollinger and Shaffran (2005) suggest training nonexperts in creativity rating before asking them to produce their final assessments. Moreover, in line with Simonton (2010), experts agree with non-experts to higher degree, on what is creative, if their assessments are made in a more pragmatic domain, which should be the case of design (compared to pure art).

By contrast, some authors have found that inter-judge agreement can be lower for experts than for nonexperts (Dollinger, Urban, & James, 2004; Hickey, 2001; Kaufman, Gentile, & Baer, 2005). Caroff and Besançon (2008) interpret this finding as evidence that expert judges are influenced by their subjective understanding of creativity, which they developed individually in the course of acquiring their own experience.

5 Criteria for assessing design creativity

5.1 Criteria used to assess creative design

It is crucial for people involved in the development of design products to have adequate and reliable means of assessing the qualities (including creativity) of these products (Besemer & O'Quin, 1999).

Several authors have tried to identify the fundamental criteria for assessing product creativity. Bearing in mind the basic definition of creativity, we know that in order to be creative, a product must be **original** and **adapted to the context** (Amabile, 1996; Lubart, 1994; Runco & Charles, 1993; Sarkar & Chakrabarti, 2008; Sternberg & Lubart, 1995). Nevertheless, authors found other observable characteristics of products that can be considered as the creativity “indicators”.

Cropley, Kaufman, and Cropley (2011) stress the fact that in design, to propose a valuable product is more important than just to come with a new idea. They employ the term *innovation* which, they point out, has much in common with creativity, quoting West's (2002) claim that “Creativity is the development of ideas, while innovation... is the application of ideas.” Cropley and Cropley (2005, 2010b) suggest that design has a special quality called *functional creativity*. They defend the functional model of creativity, in which novel products are mostly those that are useful for society and which arouse admiration for their effectiveness. Moreover, a design product must satisfy a set of predefined specifications and needs to be assessed by different metrics from those used for nonutilitarian products (Shah, Vargas-Hernandez, & Smith, 2003).

Some authors have performed some in-depth research on these two main criteria to observe their importance for creativity assessment. For instance, Caroff and Besançon, (2008) showed that creativity ratings for advertisements are higher for products with a high level of originality, but this effect is outweighed by higher appropriateness levels in creative products. As we can see, creativity is perceived of as a characteristic that can be observed to a greater or lesser degree, not as something that products either have or do not have. There are different levels and kinds of creativity (Cropley et al., 2011). For example, in their functional model of creativity, Cropley and Cropley (2005) classify creative products as follows: routine products (characterized exclusively by effectiveness) at one end and innovative products (characterized by effectiveness, but also novelty, elegance and genesis) at the other end, with increasingly original and elegant products in between. They stress the fact that routine products may be very useful and contribute to the progress of a

society even if they are not truly novel, but only adaptations of existing ideas (Sternberg, Kaufman & Pretz, 2002). On the other hand, purely aesthetic products, which have novelty but not effectiveness, are described as *pseudo-creative* by these authors.

In two subsections below, we provide an overview of existing criteria. First, we extract lists of measures from existing studies on creative design assessment, in order to find out which of criteria are the most frequently used in research. Based on these findings, in the second subsection we aim to describe each criterion on which the creativity can be assessed.

5.1.1 Overview of criteria based on existing studies

How can we state that a product is more or less creative? We propose an overview of some examples of existing studies, in which authors have suggested criteria that could be used in product creativity assessments. All the criteria are summarized in a table at the end of the subsection.

The Creative Product Semantic Scale, developed by Besemer and O'Quin (1987), is based on three criteria: *Novelty* (the product is original, surprising, and germinal), *Resolution* (the product is valuable, logical, useful, and understandable), and *Elaboration and synthesis* (the product is organic, elegant, complex, and well-crafted). These dimensions are rated on separate scales (e.g., *surprising-unsurprising*, *elegant-inelegant*), via 43 items (Besemer & O'Quin, 1999).

Horn and Salvendy (2008) looked at the criteria of product creativity (in the field of websites) from the consumer's point of view, focusing on satisfaction and purchasability. According to their Product Creativity Measurement Instrument, the main product creativity criteria are *Affect* (which refers to emotional impact of the product, measured by two sub-dimensions: arousal and pleasure), *Importance* (deals with the usefulness of the product to the customer, using two subscales: relevance and significance), and *Novelty* (considered as the most important determinant of creativity).

Cropley et al. (2011) tested the criteria proposed by other authors in design-creativity assessments. Using a Factor Analysis they obtained five most important criteria: (1) *Relevance and effectiveness*, a criterion, which includes performance, appropriateness, operability, safety, durability and correctness concerning the conventional knowledge and technique, (2) *Problematization* criterion, which allows to perceive shortcomings in other existing solutions, the possibilities to improve existing solutions or to anticipate the effects of changes, (3) *Propulsion* criterion, describing new ways of use, new approaches, new perspectives on possible solutions or extending what is known in a new direction, eventually showing the usefulness of the previously abandoned approach, (4) *Elegance*, which reflects the holder's recognition, the impression that the solution is well done, complete, environmentally friendly, well-proportioned, skillfully executed and that its elements fits together in harmonious way, (5) *Genesis*, reflecting the solution's capacity of suggesting a novel basis for further work and new ways of looking at existing problems, solving apparently unrelated problems, drawing attention to previously unnoticed problems, suggesting new norms for judgment or a new conceptualization of the issue. These criteria constitute the Revised Creative Solution Diagnostic Scale (CSDS) and can be used as observable characteristics of a product.

Table 1: Criteria to assess creativity studied by different authors, with the proposition of synthesis under common labels

Besemer and O'Quin (1987)	Dorst and Cross (2001)	Horn and Salvendy (2008)	Kreitler and Casakin (2009)	Cropley at al (2011)	Demirkan and Afacan (2012)	Proposition of a common label
				genesis		originality
				problematization		
				propulsion		
novelty	creativity	novelty	originality		novelty	
resolution	ergonomics	importance	functionality or usefulness	relevance and effectiveness		user-appropriateness
	technical aspects		fulfilling design requirements			brief-appropriateness
	business aspects		considering the context			
elaboration and synthesis			elaboration	elegance (skilful execution)	elaboration (integration, adequacy)	elaboration
		affect			affective characteristics	affect
	aesthetics		mastery of aesthetic skills	elegance (harmony of design elements)	elaboration (harmony of design elements)	aesthetics

Kreitler and Casakin (2009) identified creativity criteria specific to architectural design on the basis of a survey of 25 experienced architects. They only listed criteria that were described as important aspects in evaluating architectural design by at least 52% of respondents. These design-specific criteria (*Mastery of aesthetic skills*, *Functionality or usefulness*, *Fulfilling design requirements*, and *Considering the physical context*) were added to Guilford's (1981) factors of individual creativity (*Fluency*, *Flexibility*, *Elaboration*, and *Originality*). When they applied these indicators as criteria for assessing architectural designs, the authors found a high degree of congruity between expert judges (professional architects).

Dorst and Cross (2001) were interested in correlations between the overall quality of the design concepts and the following criteria: *Creativity*, *Aesthetics*, *Technical aspects*, *Ergonomics* and *Business aspects*. The closest correlation with the overall judgment was found for the ergonomics criterion.

Demirkan and Afacan (2012) carried out a very detailed study, conducting factor analyses and looking at the interactions between the numerous creativity indicators (based on previous product creativity studies). They focused on the importance of the *Design elements* (shape, size, proportions, geometric relationship, figure-ground relationship, and colour) and the way they were put together (harmony, rhythm, unity, variety, repetition, balance, and order). Their findings revealed that shape was both a cognitive and an affective characteristic, and that harmony, geometric relationship and figure-ground relationship enhanced the elaboration factor. Their study indicated that creative design

characteristics depend on the assessor's cognitive and affective perceptions. Moreover they concluded that, in order to be creative, a product should have the following characteristics: (1) *Novelty* (new, novel, unusual, unconventional, unique, original, infrequent, extraordinary, different, eccentric, and exciting); (2) *Elaboration* (integrated, polished, refined, adequate, deliberate, detailed, sensible, balanced, and coherent); and (3) *Affective characteristics* (appealing, delighted, good, and pleasant).

Table 1 sets out the criteria proposed by the authors quoted above. We synthesized the criteria with similar meaning under common labels, respecting the definitions given by different authors. This allows us to unify the terminology used by researchers in an inconsistent way. Consequently, we obtain overview map, which allows us a comparison of criteria proposed by different studies. We have excluded those that measure the design process (*Fluency* and *Flexibility* from the Kreitler and Casakin study), only retaining measures concerning design outcomes.

5.1.2 Definitions of chosen criteria

As we can see in the Table 1, both main criteria of the creativity definition *Appropriateness* and *Originality* are frequently cited by authors. Nevertheless, we observe also other important criteria for assessing design products. Bellow we describe more in detail each label with regard to the original definitions.

The first criterion, synthesized under the *Originality* label, appears in all the cited studies, mostly under the terms of novelty or originality, some authors (Horn & Salvendy, 2008) considering it even as the most important. It is mostly described as a characteristic that makes a product surprising, new, infrequent, unique. Dorst and Cross did not use these terms, instead they employed “creativity” as one of criteria set, and since other criteria are mostly connected with appropriateness and aesthetics, thus the creativity criterion is the only one with the originality penchant. Cropley at al. mentioned different advanced approaches to novelty, like finding fields of improvement for existing solutions (problematization), new approaches to the existing problems (genesis) or discovering new directions (propulsion).

The label *User-appropriateness* was used for all the criteria that comprised the functional and useful aspects of products. The terms used to describe this criterion are various, underlining that the creative product's use should be understandable (Besemer & O'Quin, 1987) and easy to operate or ergonomic (Cropley at al., 2011; Kreitler & Casakin, 2009, Dorst & Cross, 2001) and should be relevant to existing knowledge (Cropley at al., 2011). Nevertheless, since our research aimed the graphic area of design, we decided to use the term *Public-appropriateness*, since this concept can be taken into account for the products which are not functional in a strict sense. Such is the case of graphic design, for which the appropriateness “evokes people's intention to purchase, adopt, use, and appreciate” the creative product (Piffer, 2012).

The appropriateness term can be also approached differently, as *Brief-appropriateness*. We suggested this label to designate the degree to which the product fulfills different aspects of the brief and of the context for which it is designed, such as constraints of technical or financial nature, all the requirements described in the brief or expressed by the contractor, but also added by the designer, resulting from his or her professional knowledge.

Elaboration can be perceived as suggested in Guilford's divergent thinking tests and measured by number of details. But most of authors approached this criterion as the degree to which the product's different elements were synthesizing them in an elegant and

complete way (Besemer & O'Quin, 1987; Cropley et al., 2011). Demirkan and Afacan (2012) put the emphasis on the fact that elaboration is close to aesthetic elements, such as harmony or geometric relationship (even if in their study, elaboration is more about the detail finishing and consistency). Thus, for some authors, we put the same criterion under both, elaboration and aesthetics label. For Cropley et al., it was the case of the Elegance criterion and for Demirkan and Afacan, it was the case of the Elaboration criterion.

The *Affect* label was proposed, since two groups of authors mentioned as important the fact that a product should incite pleasure (Horn & Salvendy, 2008) as well as it should be appealing and attract public attention (Demirkan & Afacan, 2012).

Concerning the *Aesthetics* label, as we mentioned above, aesthetics can be perceived as being strongly connected with the elaboration criterion. Nevertheless the aesthetics criterion concerns less the general quality of execution and synthesis, and more the general harmony and the pleasant use of design elements.

Even if they are less quoted by authors, the two last criteria: the emotional impact of the design and its aesthetical quality, are recognized as important by many approaches, such as *emotional design* and *Kansei engineering*, which emphasize the need to take the aesthetic dimension into account, some authors have suggested that the emotions generated by the aesthetic experience are important to the overall appreciation of a given product (Norman, 2004; Yan, Huynh, Murai, & Nakamori, 2008; Bouchard et al., 2009). They can also affect the appropriateness criteria, such as the apparent usability of an interface (Tractinsky, Shoval-Katz, & Ikar, 2000).

6 Different levels of design

According to Norman (2004), design can be perceived at different levels, since human responses are determined by various factors. Some of them are external to the judge, controlled by advertising or brand image; some others come from the judge's own experience. Following studies of emotion, Norman suggested that there are three different levels of approach to design: visceral, behavioral and reflective.

The visceral level results from the genetically determined, simplest reactions to the sensory information coming from the environment. Norman gives the examples (while pointing out that the list is still under dispute) of conditions offering positive affect (e.g. warm places, sweet tastes, smiling faces, symmetrical, rounded objects) and negative affect (e.g. heights, bitter tastes, sharp objects, misshapen human bodies).

The behavioral level neither depends on appearance nor on rationality. It focuses completely on usability, function, understandability and physical feeling. This is typically important for product design, since, as summarized by Norman, "if a potato peeler doesn't actually peel potatoes, or a watch doesn't tell accurate time, then nothing else matters" (p.70).

The reflective level covers the territory of the message, of the culture and the meaning of the product. For example, when something is perceived as pretty, this judgment comes from the visceral level. Nevertheless design professionals aim to overcome this first reaction, with rationalism, since they wish their work to be recognized as creative and deep, while "pretty" is often a synonym of superficial. Reflective perception of design is influenced by the judges' self-image, their belief of "what is right" and the causes they support.

In this thesis, especially in chapter 6, we will apply Norman's model, especially the visceral and reflective approach, to the assessment of creativity of graphic design.

PART II – RESEARCH FRAMEWORK AND THREE STUDIES

Chapter 3 Research framework

1 Studying variability of creativity evaluations

Only few studies have pondered the influence of individual differences on creativity assessments. The CAT proposed by Amabile (1982) is an elegant tool for measuring creativity: it yields data with ecological validity and is neither time- nor resource-consuming. However, as its name suggests, it looks for consensus, not for sources of variability in creativity assessment. Most studies use a consensual approach in order to infer a universal means for creativity assessment, and lack of consensus between the judges is perceived as a failure. Therefore, there is a low number of research works focusing on differences in creativity judgments in creative design. Moreover, as shown by a study of literature, the CAT is not currently widely used in the design field (Jeffries, 2012).

We therefore decided to examine the extent to which judgments of creative design can be consensual and to identify the factors behind variability in the judgment of creative productions.

In the previous chapters, we examined the criteria identified by a number of experimental studies as being important to assess creative products and design. We grouped them to obtain the general categories of criteria. We now contrast these findings with empirical findings, collected from actual design professionals, during real-world situations where they are required to judge products of creative work.

Moreover, we are interested in the way in which these findings could differ depending on judges' different profiles and different viewpoints. Therefore, our study aims to analyse the divergence of viewpoints on creative design, using both (1) methods from the ethnographical domain, frequently used in ergonomics, which are the most appropriate for the purpose of collecting and analysing empirical data and (2) experimental methods,

frequently used in cognitive psychology, which are appropriate for hypothesis testing and for quantitative data analyses.

We focused our observations on two factors. First, in line with studies conducted in other areas of creativity research, we compared assessments made by different groups of judges (in the field of design, such studies are unfortunately few and far between). Second, we focused on the criteria used in design assessments and measured the degree of influence these criteria have on the assessments that judges actually make.

The studies presented in this thesis bring some new elements of knowledge, from the point of view of fundamental research, as well as from an applied perspective. Fundamental research will be provided with new empirical data on product creativity assessments in the area of graphic design that, in its current state, contains fewer studies than other areas, such as architectural design or product design. Moreover, it is essential to maintain the ecological validity of creativity measures, by collecting data about the assessment criteria within different contexts, analysed with the help of ethnographical methods and multiple feedback approaches. These methods provide the knowledge that is useful for creativity assessment methods that employ the participation of judges. It yields some insights about the nature of influence that the judges' backgrounds and level of experience can have on their assessments: which criteria guide their creativity assessments and which give more probability to obtain the highest inter-judge agreement.

From an applied perspective, our study can bring some insights to design companies, as well as individual designers, so that they could easily simulate different viewpoint on their work and better prepare arguments to defend their design choices. The results of our research could be also interesting for people, who are professionally involved in design assessment (teachers, critics, art directors), since it would allow them to understand some mechanisms of their own assessment process.

2 Research questions and general hypotheses

In the studies conducted for this thesis, we have focused on two general aspects of creative design assessment: the **assessment criteria** used for design creativity by different **profiles of judges**. Moreover, we wish to compare these factors in three different **types of assessment context**.

Criteria to assess creative design. The aim of this thesis is to understand to learn more about the dimensions used in previous research made in different fields of creativity (*Originality*, *Public-appropriateness*, *Brief-appropriateness*, *Elaboration*, *Affect* and *Aesthetics*, see Chapter 2) by validating them with empirical data retrieved within the field of graphic design and through different assessment contexts.

We wish to know which of those criteria mostly constitute the judges' mental representations of creativity. According to Tversky and Kahneman (1974), the availability of these representations would affect the way in which people use heuristics in their judgments. Therefore, it would be interesting to find out on which criteria the heuristics used to assess graphic design creativity are mostly built.

Moreover, since Norman (2004) suggests that there are different levels of approach to design, we aim to determine which criteria are decisive for assessments made based on those different levels. Furthermore, by making qualitative analyses of judges' creativity assessments, we could complete our knowledge on the real use and understanding of those criteria.

Judges' profiles. In line with Glăveanu (2010) and Wojtczuk and Bonnardel (2012), our aim is also to explore the use of criteria in creativity assessment using a multiple feedback methodology. Numerous researchers use *expert judges* to provide their creativity assessment. But the profiles of these experts vary, going from graduated design students, to professionals, or design teachers. In this thesis, we aim to analyse differences that occur within the population that could be put under the common label of “design experts”. Thus, our goal is to find out what is the influence of the judges' profile on the consensual character of their assessment of design creativity. We agree with authors who claim that creativity judgments rely on shared agreement on specific meanings within a group of people (Csikszentmihalyi, 1988; John-Steiner, 1992) and that these meanings could be different following to the social context in which specific individuals are rooted (Glăveanu, 2010). In this thesis we will focus on the context understood as professional activity – its nature and quantity – typical for profiles connected with design. We presume that this activity affects the judges' mental representations of creativity in design, which could affect the perceived or implicit importance of specific criteria within design creativity judgments.

Norman's model differentiates between the users' and designers' approaches to design (Norman, 1988). We wish to explore these differences using empirical data and to extend this approach by introducing domain “gatekeepers”, *i.e.* people who decide which products should be accepted within the domain and which should be rejected (based on Csikszentmihályi's systemic model of creativity). We chose two positions between the examples of positions occupied by gatekeepers given by Hooke et al. (2003), with a help of a professional designer and design teacher. We decided that studying profiles of design teachers on one hand and art directors on the other hand, would provide us with two different approaches: one from the educational environment, the other from the professional one.

Our goal is to study the profiles of judges corresponding to our theory-based groups, but that would also match the profiles of professional statuses existing within the design domain. In order to have a real-world, relevant approach to the professional environment of design, we selected the profiles for our studies with the help of a professional designer and design teacher. In the end, we decided to use the following profiles for our research:

- **Professional designers** (in the areas of both product design and graphic design), who carry out their creative work within the domain of design.
- As we have seen in the previous chapters, their creative process requires on the one hand a focus on the project's constraints and from the other hand, an enlargement of the area of research for creative ideas (Bonnardel, 2000, 2006). We expect that these activities can influence their judgments of design. Managing constraints consists especially in attributing them different levels of importance, depending on the designers' evaluative referents (Bonnardel & Sumner, 1996). This could result, in line with the findings of Lera (1981), in the **low inter-judge agreement** in designers' assessments made **on the Brief-appropriateness criterion**, since judges may assign a high level of importance to different constraints of the design solution. Concerning the search for new ideas, it encourages designers to expand their search to a large quantity of uncensored and unexpected ideas, which would result in an approach that is less critical and more **focussed on Originality**. Moreover, designers use multiple perspectives and take into account the viewpoints of different customers and users (Bonnardel & Sumner, 1996), thus we expect that

in our study, this should lead to the **use of multiple criteria**, with a special focus on **user-appropriateness**..

- **Design teachers** – Domain gatekeepers share the domain knowledge and their role is to focus on the student's learning process and individual progress (Birenbaum, 2007). This could affect their way of assessing product creativity. We expect their assessments to be not only focused on product creativity, but also to induce some judgments connected with process- and person-creativity.

As suggested by Hooke et al. (2003), gatekeepers transmit the *domain-relevant knowledge*, which they should make them able to identify the required degree of appropriateness and novelty. On this basis, we expect that in the area of design assessment this group should reach a relatively high inter-judge agreement on the ***Brief-appropriateness*** criterion (especially, because the teacher is the one who explains the brief to his or her students) and ***Originality***. Nevertheless, we should remember that teachers, in their professional activities often apply formalized, **multiple-criteria** assessments, such as those from the Student Product Assessment Form (Reis & Renzulli, 1991), which can lead to a system of criteria in which most of them are of equal importance.

- **Art directors** – another type of gatekeepers, with more commercial profiles, since they control the release of products on the market and represent the knowledge about existing trends and about user needs.

They have access to a large amount of domain knowledge concerning facts, circumstances and issues on the area of design (Amabile, 1996) and this knowledge allows them to identify the required degree of actual **aesthetic trends** in order to be able to define the visual identity for their customers. Moreover, they should take into account other considerations that are important for their audience, such as religious, political, economic or ethical considerations. In consequence, the ***Audience-appropriateness*** should also be a high-priority criterion.

- The last group would be called the **audience** while comparing different backgrounds, we will call this group the audience, but during the comparison of experience levels, this group will be called **laypeople**. Participants from this group are people who are neither designers, nor design experts, but who are potential users (for product design), or the potential audience (in the case of graphic works) of design work.

As described by Chan and Chan (1999), this group builds implicit theories about the creativity, based on their belief system about design and on self-related interpretations. During assessments, it might be the source of lower inter-judge agreement, compared with experts, whose knowledge about design is based on explicit theories (based on domain knowledge established by researchers and specialists in the field). We assume that because of their subjective understanding of design attributes, the users or audience would use various criteria to assess design, thus **their inter-judge agreement would be lower** than in other groups, which would be in line with results of such researchers as Kaufman et al. (2008), Lee et al. (2005) or Plucker et al. (2008). Moreover, following to the model proposed by Zeng et al. (2009), users rely mainly on process information on design with their senses, which in the first place would influence their ***Affect***. Thus, we expect that their assessment of design creativity is, in the first place, based on their emotional reaction to the assessed design. Only next does the comparison come

between the design's attributes and the creativity criteria (which, as we have already pointed out, are not founded on the domain knowledge base).

In line with many studies that analyse the impact of the judges' experience on their assessment of creative artefacts (Caroff & Besançon 2008; Hickey, 2001; Kaufman et al., 2009; Kaufman et al., 2008), we included the *level of professional experience* factor within some of our studies. The previous studies focussed especially on the differences between inter-judge agreement within the groups of experts and groups without any experience in the given domain. In this thesis, we focus not only on inter-judge agreement, but also on the differences concerning criteria used for assessing creativity.

We chose 10 years of professional experience as limit differentiating experts from other participants, as suggested by a number of authors (Bloom, 1985; Gardener, 1993; Hayes, 1989). We compared data obtained from these participants with those obtained from less experienced design professionals and from laypeople. Therefore, when analysing the results, we divided our samples into three groups:

- **Asserted experts** - participants with more than 10 years of professional experience within the domain of design,
- **Intermediary experts** - participants with between 5 and 10 years of professional experience in design,
- **Laypeople** – participants with no professional experience in the design at all.

Assessment context. We consider that the assessment process can depend not only on judges' individual differences, but also on the type of procedures comprised within the experimental methods used to collect the assessments. Moreover, studies in the field of individual differences often require large amounts of data in order to validate significant differences between groups. Since it is not easy to obtain large quantities of data within a population with specific professional profiles such as those chosen for our studies, we consider that studying our variables using different methodologies and different contexts would give a more precise and ecologically valid picture of the criteria used for creative design assessment.

We conducted three types of studies:

1. In the first type of study, we collected the **criteria considered as important** to assess creativity in design, focusing on the representation and mental representations that people have of creative design assessment, not on the degree to which they adhere to a pre-set list of existing criteria.
2. In the second type of study, we aimed to find out how judges use the **criteria imposed by the experimenter** and whether their assessments were consensual.
3. In the third type of study, we observed which **criteria** were **spontaneously expressed in a real time design assessing situation**, when they can freely express all their thoughts about the assessed artefacts.

Hypotheses. The objectives of this thesis can be presented within two general hypotheses:

- We presume that the evaluative referents and criteria used to assess creativity in graphic design will differ in their nature and the frequency of use from those

presented in studies on product design. Our goal is to study the specificity of criteria used in this specific area.

- We expect that the judges' professional backgrounds and their level of professional experience in design will influence the nature of their creativity assessment of graphic design. More specifically, we believe that these differences can be found within the specific criteria considered the most during creativity assessments, as well as within the degree of inter-judge agreement on those criteria.

In this thesis, our goal was to identify which criteria influence the most the creativity assessments of specific judges' profiles. Thus, we used different assessment contexts, in order to find out which criteria are used in a most salient way. These different contexts allow us to compare the criteria based on participants' mental representations of creativity, used to attribute scores to creative outcomes and spontaneously verbalized during real situation of design assessment.

3 Methodologies

3.1 Data collection

Depending on the procedures used for the assessment, different methods of data collection and analysis are required. In our studies, we used three types of data collection methods:

- Surveys,
- Experimental methods,
- Semi-structured interviews.

All of these methods are described in greater detail in the separate chapters concerning each study. Nevertheless we present here a short preview in order to explain our general methodology.

The first study aims to collect the criteria spontaneously suggested by participants as the most important criteria to assess creativity in graphic design. Thus, data were collected through a **survey** with open questions. We wished to find out whether there is a set of general criteria adhered to by most people. Moreover we looked for differences between the criteria proposed by judges with different professional viewpoints and with different levels of expertise in design. We examined two areas of design – graphic design and product design – in order to compare them and to find out the common characteristics that would help us to identify the assessment criteria that are specific to these areas. In the first stage of this study, we will collect and analyse qualitative data, which will be transformed into quantitative data and compared using statistical methods.

The second study aims to analyse the scores attributed by judges to graphic designs, with regard to predefined criteria. Our aim is to determine the degree to which the judges' assessments are consensual (1) within the entire sample, (2) within groups of judges with different profiles. For this study we used an **experimental method**, using quantitative data provided by scores that judges assigned on specific criteria.

The third study focuses on the criteria expressed by participants in real time during their open verbalisations while assessing designs. Data were collected during **semi-structured interviews**. We were interested in judges' spontaneous comments and we were looking for differences and similarities between the approaches of our participants, still with regard to their professional viewpoints and level of expertise in design. As for the first study, our

approach comprises two stages: first, collecting and analysing qualitative data; and second, transforming them into quantitative data to use statistical models of comparison.

3.2 Data analysis

Qualitative analyses. In the first and the third studies, to analyse qualitative data, we adapted *Grounded Theory*, the methodology used in ethnographical sciences to analyse the content of documents collected by empirical research.

Grounded theory (Glaser & Strauss, 1967) is a research method that operates through data analysis as a basis for building hypothesis. This is opposite to a traditional experimental method, which begins with a hypothesis and organises the data in order to validate or to refute it. This method allows the conceptualisation of data included in all kinds of transcriptions or verbalizations, by merging them into new concepts. Goulding (2002) describes grounded theory as a research methodology best suited for researchers whose main goal is to build the theory from the ground, not from previously proposed hypotheses. The method of data analysis offers clear stages of analytical activities:

- *Open coding* includes composing a set of categories that form the conceptual elements of the research problem. In this stage, key factors and associated concepts are extracted from the analysed interviews. Therefore, the main codes are identified and those with similar content are grouped together by three specialists. In consequence, some key concepts are distinguished to tackle the research problem.
- *Axial coding* focuses on organizing, reducing and clustering the conceptual categories generated from the open coding.
- *Selective coding* comprises the shift of attention towards one or more central categories, the detailed development of these categories, and the selection of core categories. This gives some direction to theory development.

In our case, we use this method to analyse the verbal utterances and responses to the open questions that are sources of qualitative data in the first and the third studies. Our goal is (a) to proceed with analysing these data independently from existing research, and (b) only after this, to compare the findings of other authors.

To avoid confusion and be more precise about the terms used in this study we used two different terms to talk about criteria.

- The term *evaluative referent* is used to indicate a factor taken into account by a judge in order to assess design creativity (Bonnardel, 1996). More specifically, we use this term in the context of participants' verbalisations.
- The term *criterion* is used for categories assembling similar evaluative referents. Moreover, we use the term *criterion* for assessment tasks in which judges have to assign scores from 1 to 7 – the scores are assigned on specific *criteria*.

Moreover, in the first stage of both studies, we apply data analysis methods from grounded theory methodology. During this stage, we analyse and categorise the concepts representing the judges' evaluative referents (ERs) into criteria. The second phase consists in analysing the findings from the first stage with statistical methods, which requires quantitative data. In order to obtain them, we adapted the method of grounded theory to our needs, by taking into account the number of occurrences for each ER or criterion.

Our method was organized in following way:

- 1) When analysing each response of a participant, we identify the main codes that play the evaluative role within the assessments in order to group them within categories of codes with similar meaning.
- 2) The names given to the categories aim to be the most representative for the grouped codes and at the same time to express *evaluative referents for creativity assessment*. A list of evaluative referents can thus be produced, with each referent accompanied with the original quotations from the participants' verbalizations and data containing information about the assessor (his or her professional profile and level of expertise).
- 3) Finally, the obtained criteria are compared with the results of other research works, in order to find common *criteria for creativity assessment*.

For example, when a judge writes or says that he or she appreciates the choices of symbols used in a graphic production, we consider it as an evaluative referent. Moreover, we assume that the code *choice of symbols* is similar to *meaning of the picture* (used already by other participants') and we give to our evaluative referent a common label: *relevance with the subject*. This allows us to know in details which evaluative referents are used by participants to spontaneously evaluate design and we can compare the number of their occurrences.

We categorise evaluative elements within criteria, in order to obtain more general rules of assessing design. To do so, we take into account the existing state of art in creativity assessment. In the case of our example, *relevance with the subject* is close to the measures existing in the literature, such as the measure used by Kreitler and Casakin (2009) that indicates if the artefact is fulfilling some previously established requirements. Thus, we can group it with other evaluative referents linked with requirements (*Respecting codes of the domain, Practical constraints, Values*) under the single criterion of *Brief-appropriateness*. This allows us to compare our findings with those of other authors, but also, we chose to use *Brief-appropriateness* as one of measures of creativity on which judges can assign scores, while assessing the specific products.

During these analyses, the information about participants who mentioned each evaluative referent or criterion (their number and characteristics concerning professional background and level of experience) are retained. This method allows us to use the obtained results as variables for statistical analysis in the second stage of the study, which aims to understand the differences in the approach to the creative design assessment, between different groups of participants.

Quantitative analyses

In order to select the most important evaluative referents (ERs) and criteria from the survey, we used the **thresholds of 50% and of 30%** of judges that agreed about its importance for creativity. Since the participants did not choose the criteria from an imposed list, but retrieved them from their mental representations built from personal or professional experience with creativity – the final list of criteria was too long to proceed with any other statistics.

A repeated-measures analysis of variance (**ANOVA**), as being the most appropriate way to analyse mean differences between groups (categorical variables), taking into account a within subject variable that have as many modalities as there are opportunities of measure

To verify the consensual character of assessments, we used the **Cronbach's Alpha**, commonly used to estimate the internal consistency for sample of participants. Amabile (1996) justifies the use of Cronbach's alpha coefficient by the fact that it is a static indicator generally used to valid the internal reliability of questionnaires and psychometric tests, while the estimated degree of agreement can be understood as reliability, since it estimates the agreement of the same set of judges on other products.

To find out which criteria contributed the most to the overall creativity assessments, we used the **linear regression**.

In order to find out which groups of judges seem to proceed similarly within their creativity assessments, we examined the **Multiple Factor Analysis** of the criteria distribution across the different judges' profiles.

In order to analyse the differences in the frequency of criteria use, appearing between groups of judges, we proceeded with **Correspondence Analysis** and the **z-scores** that expresses how far a value is from the population mean, and expresses this difference in terms of the number of standard deviations by which it differs (Kirkwood & Sterne, 2003). This analysis allows us to know which criteria are used by a group of judges with a specifically high or specifically low frequency, compared (1) to other criteria used by this group and (2) to other groups.

Chapter 4 – Criteria declared as important for the design creativity assessment

The aim of our first study is to collect the criteria that people consider as important when assessing creativity in design. In this study we aim to collect criteria reflecting the theoretical viewpoints, mental representations and declared priorities of our participants. In order to obtain that information, we asked the participants to express their criteria independently of any specific examples of design. Moreover, the participants involved in this study were not provided with lists of suggested criteria, but had to quote spontaneously what is important to them, while assessing the creativity of design.

The objective of this study is threefold.

First, we wish to find out whether there is a set of general criteria adhered to by most people.

Second, we wish to compare criteria proposed within two areas of design: graphic design and product design.

Finally, we aim to analyse differences and common points between criteria proposed by participants with different professional viewpoints and with different levels of expertise in design.

1 Method

For this study we choose to use a survey for collecting the qualitative data, which were then analysed with the use of an adapted version of the grounded theory method (see Chapter 3). The findings were then transformed into quantitative variables and statistically compared, in order to determine the differences observed between different groups of participants.

1.1 A survey with open questions

The most appropriate way to discover what is important for participants to assess creative design is simply to ask them about the criteria they think are the most suitable for this domain. Therefore, for the phase of data collecting we used an online survey with open questions.

When choosing the method for data collection, we have to measure its advantages and weaknesses in order to select the most suitable one. Our survey method has some advantages:

It allows us to reach a large number of participants, which is especially interesting if our goal is to reach people with very specific professional profiles. Participants can express themselves freely, since they remain anonymous and are not influenced by the experimenter's presence.

It is easier for the researcher to process the results, when the data is already pre-organized by the existence of specific questions and specific response formats.

Of course, this method has also several drawbacks:

- Participants mention the criteria without the evaluation context, which can result in very theoretical responses. Indeed, participants could express criteria that they would not apply in a real assessment situation.
- It is impossible to know if the results obtained in this way reflect the participants' knowledge about what should be important (e.g. they could quote authorities in the field or learned theories) or their own personal values.
- The open character of the questions leads to a risk that some criteria would not be mentioned because they were simply forgotten or seemed too evident to be mentioned.
- The free format of responses results in the variety of used vocabulary, thus some participants can express their criteria in ambiguous ways.

Nevertheless, some of these weaknesses can be transformed into advantages: this method provides a picture of specific mental representations of design creativity, expressed by participants with different professional viewpoints and levels of experience in design.

The criteria that illustrate extremely personal visions are detected by the simple fact that they are personal, thus quoted by an extremely low percentage of participants and rejected during the data analysis. Ambiguous elements are also eliminated, in order to avoid false interpretations. Data analysis also solves the problem of differences in vocabulary, even though qualitative data always leaves some possibilities of mistakes in interpretation.

1.2 Data analysis

To analyse the responses given during the survey, we adapted the grounded theory method described partially in the previous chapter. More specifically, in this study we proceed as follows:

- First we identify **codes**, by collecting terms used by participants of the survey to describe the criteria of creative design assessment. Thus we obtain a list of terms, which comprises different propositions of criteria expressed in various ways and by different numbers of occurrences.
- Second, we group the terms with similar content, in order to find the common concepts. Therefore, we assembled them in order to create the categories, which we treat as **evaluative referents** (ER) – entities spontaneously proposed by participants as important to evaluate creativity (Bonnardel, 1996). For each category, we find the most representative label.
- Third, these ER can be compared with findings resulting from other research (quoted within the state of art) in order to obtain the final **criteria**.

Using this methodology, we expected to find relatively detailed information about the representations and approaches to creative design assessment. More precisely, (1) our data show the vocabulary spontaneously used by people to describe criteria for assessing design in two areas (graphic design and product design), (2) we present a list of ER resulting from grouping the participants' suggestions and compare our findings with those made by other authors, (3) we compare the quantity and the nature of ER and of criteria mentioned by participants with different professional backgrounds and levels of expertise, (4) we compare the results within both graphic and product design.

2 Participants

2.1 Recruitment

Our participants were recruited via email invitation to fill in an online questionnaire. We sent invitations to participate in our study to approximately 380 design professionals and 60 laypeople. 43 design professionals (11%) and 20 laypeople (33%) responded positively to our demand and filled correctly the questionnaire.

This way of collecting responses seems the most appropriate, as it allows us to reach design professionals in the entire country. We could access a high number of people with the appropriate profiles, even if we knew that the percentage of responses would be lower than if we contact them in person.

2.2 Establishing specialization profiles

We asked each participant what was precisely his or her professional activity and to indicate the number of years spent on this activity. They could choose one or more activities between:

- Graphic or product designer
- Art director
- Design teacher
- Other (please, specify your professional activity)

We reserved the other category for participants whose activity was not connected with design, in order to obtain the control group of laypeople. We needed this group to compare the results obtained by the design professionals with those obtained by participants having no professional experience in design. Moreover, this group can be considered as the design users (in the case of product design) or the design audience (in the case of graphic design) and their opinion reveals the viewpoint of users' experience and their implicit theories about creativity (Chan & Chan, 1999).

Numerous participants indicated multiple professional activities, thus we decided to describe our sample in a more detailed way. Many professional designers are also art directors. Because of this, we decided that being an art director is a decisive criterion, since it changes the viewpoint on the domain: art directors are supervising the global directions of design and often they give directions to designers. We therefore separated our participants who were only designers from those who were art directors, even if their activities were multiple.

Being a design teacher is an even more decisive criterion, as being able to explain the design to others, often from a theoretical perspective, also changes the viewpoint on the domain. If participants indicated that they were design teachers, we included them in the group of teachers, regardless of the other activities that they declared. Thus, even if a participant was not only a teacher, but also an art director or a designer, we qualified him or her as a teacher.

Because of this, for the participants professionally connected with design, we obtained the following groups:

- Designers
- Art directors
- Design teachers

2.3 Establishing experience profiles

Our aim was also to analyse differences related not only to the design specialization but also to the participants' level of professional experience in design. Within the same sample, we created additional groups.

Moreover, we aim to identify differences between participants, depending on their level of experience in the design field. We choose the limit of 10 years following the findings presented in the Chapter 2 (Bloom, 1985; Gardener, 1993; Hayes, 1989; Kaufman, 2007). We wished to divide participants into two groups: a group of asserted experts, who have more than 10 years of professional experience in design and a group of intermediary experts, who had no more than 10 years of professional experience in the domain. Nevertheless, the difference seems to be very large between a designer with a very small amount of experience (e.g. 1 year) and one who is close to becoming an asserted expert (e.g. 10 years). To avoid these important differences, we decided that our intermediary experts should have between 5 and 10 years of experience in the domain. Therefore, in the invitations to the questionnaire, we specified that we looked for the responses from designers having already some professional experience, thus we asked to participate only those who had at least 5 years of professional experience.

We created the following groups:

- Participants with more than 10 years of professional experience in the design domain, that we will name **asserted experts**.
- Participants with 5 to 10 years of professional experience in the design domain, that we will name **intermediary experts**.
- Participants with no professional experience in design at all, that we will laypeople.

2.4 Final sample

Finally we obtained a sample of 63 participants.

In order to obtain the groups with different professional backgrounds, we created the following groups:

- 16 **designers** who are nor teachers, neither art directors (14 males and 5 females, average experience = 9.9 years)
- 17 **art directors** who are not teachers (11 males and 5 females, average professional experience = 13.1 years. 9 of them are asserted experts and 8 are intermediary experts)
- 10 **design teachers** (9 males and 1 female, average professional experience = 15.4 years. 8 of them are asserted experts and 2 are intermediary experts)

To obtain data about the degree of experience in design, within the same population, we created the following groups:

- 21 **asserted experts** (15 males, 6 females, average professional experience = 18,3 years. The group includes 4 designers, 9 art directors and 8 design teachers)
- 22 **intermediary experts** (18 males, 4 females, average professional experience = 6,7 years. The group includes 12 designers, 8 art directors and 2 design teachers)

- 20 **laypeople (control group)** who do not work within the domain of design, having various ages and various occupations (14 males and 6 females, with average age = 30,7)

3 Material

We created 2 questionnaires using Google documents: one for designers and one for non-designers (see annex 1). Both of them comprised a short presentation of the researcher and of the study, specifying the required profile and the approximate time necessary to realise the required task. Both of them required also some information about the participant (email address and gender). Additionally designers were asked to give information about their specialization and number of years of professional experience, while non-designers were asked to give their age.

Furthermore, all participants were provided with instruction to write down, using their own words, the criteria that allow them to evaluate the creativity (1) in the graphical design task and (2) in the product design task.

4 Procedure for criteria finding

To analyse the data from all the participants' responses, we proceeded in three steps.

4.1 First step: identifying codes

We made two lists of all the terms proposed by participants as criteria to assess design. Therefore, these terms contains information about the subjective representations of what creativity in design is, and about the vocabulary used by participants to evoke the criteria to assess this creativity. The same term could be proposed by different participants.

In the first list, we collected the terms proposed to evaluate the creativity of graphical design (GD list). In the second one we collected the terms proposed to evaluate the creativity of product design (PD list).

4.2 Second step: categorisation and ER finding

Having both lists of terms, we then proceeded to categorize their contents, in order to find those, which could be interpreted as synonyms and label these terms under the same concept. To reduce the subjective character of the procedure, three participants were involved in this task: two researchers in the field of creativity and one professional designer.

The participants grouped the terms into categories and indicated how to name each category so it represents a criterion to assess creative design. We use the term ER to indicate a principle, by which the creativity in design may be judged, the label of each ER being chosen to group the terms proposed by different participants, having close meaning.

For example, a label *Innovation* was suggested to group the following terms from the GD list:

originality, presenting the message in an original way, original processing, new, innovative, surprising, cessation of what is usual, unusual, different, not ordinary, new idea

The same label was suggested to group the terms from the PD list, where we could find several similarities, but also some differences with the items from the GD list:

originality, novelty, surprise, novelty of the concept, impression of novelty, innovation, unknown, exploration of what is seldom used in actual design, cessation with existent codes, cessation, difference with other products of the same type, innovative technological contribution, new shapes, new possibilities, contribution to the way of use

We eliminated the items that appeared just once, without being linked to any category, or that were difficult to interpret.

To understand which ERs are the most commonly mentioned, we analysed the number of occurrences of each of them in the participants' proposals. If participants used more than one item belonging to the same ER, we considered that they used synonyms to describe the same criterion. For example if one participant mentioned *originality* and *new* it was considered as a single occurrence of the *Innovation* ER. To analyse the quantitative data, we used the number of occurrences of the ERs.

Our goal was to collect the ERs that were mentioned by at least 50% of participants, qualifying them as those of *high importance*. We also noted the ERs mentioned by 30 to 50% of participants, qualifying them as those of *moderated importance*.

4.3 Third step: identification of criteria in creative assessment

The ERs obtained in this way were examined in comparison to the criteria found within the existing state of art. In order to obtain data that were more general and more comparable with existing research, we linked each ER with one of the following criteria: *Originality, Brief- and Public-appropriateness, Elaboration, Affect and Aesthetics* (see Chapter 1).

We carried out the categorization of our ER with the goal of obtaining criteria and their number of occurrences. We then analysed the number of occurrences of each criterion in the participants' proposals. If participants used more than one ER belonging to the same criterion, we treated it as a single occurrence of that criterion.

Again, we collect the criteria of high importance (mentioned by at least 50% participants). We also marked the criteria of moderate importance (mentioned by between 30 and 50% of participants).

5 Categorisation results and Evaluative Referents finding

During the categorization process, 21 categories were created to represent 21 ER for the assessment of design creativity. We present all of them in the table 2, each with a short description and the percentage of participants that mentioned each criterion as being important when assessing creativity in product design and graphic design.

Table 2: The percentages of ERs related with creativity that were cited by all the participants in the area of GD and PD

Criteria of **high importance** (mentioned by >50% of participants)

Criteria of **moderate importance** (mentioned by between 30% and 50% of participants)

Evaluative Referent	Definition	% of participants that mentioned the criterion	
		GD	PD
Innovation	Reference to originality, uncommonness	68%	56%
Relevance to the subject	Importance of the semantic connection between the content and the theme of the graphic design	38%	
Comprehension of message	Degree to which the idea represented in the design is easy to understand for the public	29%	
Emotions conveyed	Evoking emotional reaction	24%	19%
Ergonomics and user-appopr.	Connected with usability of the design	22%	63%
Design elements	Visual components of design	22%	33%
Respecting the codes of the domain	Field codes, existing trends and other elements facilitating the recognition	21%	11%
Beauty	The appearance of design	21%	29%
Harmony	The way in which the design elements should be adjusted with each other to create an impression of unity	21%	13%
Quality of execution	Precision and finishing of the design work	17%	16%
Layout	Organization of the design elements on the given surface	14%	
Creative-person qualities	Describes the characteristics considered as important for person creativity	13%	19%
Style	Underlines the importance of some strong, recognizable traits	13%	5%
Simplicity	Use of simple means and minimum of elements	11%	13%
Appeal	Attracting the target attention	11%	8%
Creative-process characteristics	Divergent thinking and visibility of the creative process within the final production	6%	6%
Concept	Quality and the elaboration of the idea on which the design is based	5%	6%
Tools	Technology used during the creation process	5%	6%

Comprehension of use	Degree to which it is comprehensive for the user to understand the design functionalities		29%
Practical constraints	Constraints related with life of product, production process and marketing domains important to take into account within a design product		16%
Values	Moral constraints to be taken into account within a design product		5%

We can observe that for GD *Innovation* obviously has the highest number of occurrences. This was the only ER that had more than 50% of occurrences and of which we can assume that, according to the participants, it has a high importance for creativity. There was also a single ER mentioned by between 30 and 50% of the sample: *Relevance to the subject*, thus, it showed a moderate importance for participants. Probably, other ER are too specific to have higher degree of occurrence, and we could say more about the distribution of their occurrences after the second categorisation and identification of more general criteria (part 6. of this chapter).

For PD, two ERs were mentioned by more than 50% participants: *Ergonomics and user-appropriateness* has the highest number of occurrences and it is closely followed by *Innovation*. Moreover, 33% participants found that *Design elements* is also important for PD creativity.

5.1 Similarities and differences between mental representations of the GD and PD creativity

We proceeded to a comparison of differences occurring between GD and PD lists by analysing the ERs labels and the codes grouped within these ERs. We noted that even if our participants described the two areas of design finding numerous common points, their mental representations of the GD and PD creativity contain also several differences.

Innovation, *Creative-person qualities*, *Creative-process characteristics*, *Concept*, *Respecting the codes of the domain*, *Beauty*, *Style*, *Simplicity* and *Tools* are present in the PD and GD lists, although some collected codes did reveal different approaches to these two domains. For example, *Innovation* in GD includes some specific propositions, like the originality of ideas' or the originality of message presentation, while in PD we can find innovations related to technology or product use). Similarly, *Style* contains some specific characteristics for GD, since it seems to be connected with identity and drawing style, while for PD, *Style* is just a personalised form of expression.

We observed slight differences concerning the *Emotions conveyed* and *Appeal* criteria. The pleasure conveyed by the PD is specifically related to the use and the usefulness of the product. Moreover, a PD should be appealing not only by *catching the eye* (as in GD), but also by *evoking lust* and *making other people jealous*

In the case of *Ergonomics and user-appropriateness*, for PD the physical use are highlighted (*usefulness*, *usability*, *manipulation*, *comfort*). Moreover the PD *Quality of execution* compared to GD, has some extra items related with production process (*quality of manufacturing* and *quality of materials*).

Design elements and *Harmony* have also some specific ERs. Both fields have *colours* and *shapes* as basic design elements, but besides, GD also focuses on *fonts*, *images* and *contrasts*, while PD elements contain *shape*, *materials*, *colours*, *line* and *touch*. In consequence, the *Harmony* criterion concerns the harmony existing between the field-specific design elements.

The *Relevance to the subject* and *Layout* are specific to GD since the main goal of product design is rather representing a use than a specific subject and the use of layout is specific to two dimensional productions. Instead, *Practical constraints* and *Values* seem to be specific to PD. The *Comprehension* criterion varies between the two domains, in the case of PD we have *Comprehension of use* and in the case of GD, *Comprehension of message* seems more suitable.

In order to compare the results representing the participants' mental representations of creativity with those representing their ERs applied during the assessments of real designs (study described in Chapter 6), we analysed more in detail the distribution of ERs issued from the present study. We carried out two analyses of the found ER: (1) depending on the participants' professional backgrounds, (2) depending on their level of experience in design. Thus, we could understand which criteria and dimensions are the most representative for participants from different groups.

5.2 ER depending on the participants' professional backgrounds

When analysing data obtained within the groups with different backgrounds, we observed that for all the groups *Innovation* seems very important for the GD creativity (cited by more than 50% of each group) as well as *Ergonomics and user-appropriateness* and *Innovation* seem crucial for PD creativity (with a small difference for teachers that we will describe further), which is in line with the general results. Nevertheless, if we consider the preferences of each group separately, we find out that results comprise some differences.

Designers (as described in table 3) shared at least moderate interest to a higher number of ERs for both design areas, compared to other groups of participants. This is the only group that shared (moderately) their consideration for *Comprehension of message*, *Emotions conveyed* and *Respecting the codes of the domain* (in the area of GD) as well as for *Comprehension of use* and *Simplicity* (in the PD area).

Table 3: The most cited Evaluative Referents within the designers group

Criteria of **high importance** (mentioned by >50% of participants)

Criteria of **moderate importance** (mentioned by between 30% and 50% of participants)

Designers - GD		Designers - PD	
Innovation	63%	Ergonomics and user-appropriateness	75%
Comprehension of message	38%	Innovation	63%
Relevance to the subject	38%	Comprehension of use	50%
Emotions conveyed	31%	Design elements	44%
Respecting the codes of the domain	31%	Simplicity	31%

Art Directors' preferences (table 4) have exactly the same hierarchy as the general results with no additional ERs.

Table 4: The most cited Evaluative Referents within the art directors group

Criteria of **high importance** (mentioned by >50% of participants)

Criteria of **moderate importance** (mentioned by between 30% and 50% of participants)

Art Directors - GD		Art Directors - PD	
Innovation	65%	Ergonomics and user-appropriateness	71%
Relevance to the subject	47%	Innovation	53%
		Design elements	35%

Teachers (table 5) shared their consideration to *Relevance with the subject* as highly important for GD creativity, to a higher degree than other groups. Furthermore, it was the only group that considered the *Creative-person qualities* as highly important for PD creativity, while *Innovation* was considered only as moderately important for creativity. We did not want to keep the ERs related to the creative-process or –person, nevertheless in this case the number of participants with this specific profile that mentioned this ER seemed high and we judged it interesting to note it as a specificity of teachers' group.

Table 5: The most cited Evaluative Referents within the teachers group

Criteria of **high importance** (mentioned by >50% of participants)

Criteria of **moderate importance** (mentioned by between 30% and 50% of participants)

Teachers - GD		Teachers - PD	
Innovation	70%	Ergonomics and user-appropriateness	60%
Relevance to the subject	60%	Creative-person qualities	60%
		Innovation	40%

5.3 ER depending on the participants' level of experience in design

When comparing the results of participants depending on their level of experience, we observe that the distribution of the most cited ERs is not equal. This inequality is more visible for graphic design area.

Asserted experts show a common preference for a quite moderated quantity of ERs, but most of them reach the threshold of high importance. The distribution of their most shared ERs reflects the general preferences.

Table 6: The most cited Evaluative Referents within the experts group

Criteria of **high importance** (mentioned by >50% of participants)

Criteria of **moderate importance** (mentioned by between 30% and 50% of participants)

Asserted experts - GD		Asserted experts - PD	
Innovation	76%	Ergonomics and user-appropriateness	57%
Relevance to the subject	52%	Innovation	52%
		Design elements	38%

For intermediary experts the distribution of ERs is different, especially for GD: a relatively high number of ERs is considered important for creativity, but for most of them only to the moderate level. For PD, rather than *Design elements* appearing in the asserted experts' preferences, we find *Comprehension of use*.

Table 7: The most cited Evaluative Referents within the intermediary experts group

Criteria of **high importance** (mentioned by >50% of participants)

Criteria of **moderate importance** (mentioned by between 30% and 50% of participants)

Intermediary experts - GD		Intermediary experts - PD	
Innovation	55%	Ergonomics and user-appropriateness	73%
Comprehension of message	41%	Innovation	41%
Relevance to the subject	41%	Comprehension of use	32%
Emotions conveyed	32%		
Ergonomics and user-appropriateness	32%		

The only belief shared by more than 50% of laypeople is that *Innovation* has a high importance for creativity in both areas (in the GD area it is even the only ER shared by more than 30% of this group). Concerning PD area, this group of people is the only one sharing the consideration for *Innovation* to the higher level than for *Ergonomics and user-appropriateness*. In comparison with more experienced participants, laypeople give a moderate importance to *Beauty* as important for PD creativity.

Table 8: The most cited Evaluative Referents within the laypeople group

Criteria of **high importance** (mentioned by >50% of participants)

Criteria of **moderate importance** (mentioned by between 30% and 50% of participants)

Laypeople - GD		Laypeople - PD	
Innovation	75%	Innovation	60%
		Ergonomics and user-appropriateness	50%
		Beauty	40%

6 Identification of criteria in creative assessment

Our goal was to compare our findings with those of other authors. As presented in the Chapter 1, research on creativity pointed out the importance of such criteria as *Originality*, *Public- and Brief-appropriateness*, *Aesthetics*, *Affect* and *Elaboration*. The ERs revealed by our survey are more detailed; nevertheless, the aim of this study is not only to know the details of the individual approach to design creativity, but also to understand how these details can be generalised and compared to the existing state of art. These generalised categories would be especially interesting in the context of comparing the mental representations of participants with different profiles, since it allows us to enrich the existing research with some data that are new, but have the similar nature to the existing ones. We realised that the ERs, despite their detailed character, correspond to the categories represented by criteria cited in the literature. In table 9, we identify the ERs matching each criterion in both design areas (GD and PD).

This final categorisation allows us to have the information about the part that each of these criteria takes in the mental representations of creativity, depending on participants' professional viewpoint and the level of expertise in design.

Table 9: Categorisation of ERs related to GD and PD areas, with linking to the creative design criteria from literature

Criteria	Originality	Brief-appopr.	Public-appopr.	Aesthetics	Affect	Elaboration
ER for GD	- Concept - Innovation - Style	- Relevance to the subject - Respecting the codes of the domain	- Ergonomics and user-appopr. - Comprehen. of the message	- Design elements - Harmony - Simplicity - Beauty - Layout	- Appeal - Emotions conveyed	- Quality of execution - Tools
ER for PD	- Concept - Innovation - Style	- Relevance to the subject - Respecting the codes of the domain - Practical constraints - Values	- Ergonomics and user-appopr. - Comprehen. of use	- Design elements - Harmony - Simplicity - Beauty	- Appeal - Emotions conveyed	- Quality of execution - Tools

We summed up the number of times each criterion was mentioned by participants, in order to find out which of criteria were considered as the ones of highest importance for creativity (were mentioned by more than 50% of the whole sample). The results are illustrated in table 10. We found out that the criterion that was mentioned by the highest amount of participants was *Originality* in the area of GD and *Public-appropriateness* in the area of PD. Next come, for GD *Aesthetics* and *Brief-appropriateness* (eventually the *Public-appropriateness*, which was mentioned as moderately important for GD creativity) and for PD *Aesthetics* and *Originality*.

Affect and *Elaboration* did not attain the threshold of 31% in either of the two areas of design.

Table 10: The most cited criteria within the whole sample

Criteria of **high importance** (mentioned by >50% of participants)

Criteria of **moderate importance** (mentioned by between 30% and 50% of participants)

GD		PD	
Originality	71%	Public-appropriateness	73%
Aesthetics	54%	Aesthetics	63%
Brief-appropriateness	54%	Originality	59%
Public-appropriateness	41%	Brief-appropriateness	29%
Affect	30%	Affect	22%
Elaboration	22%	Elaboration	21%

6.1 Criteria depending on the participants' professional backgrounds

The analyses of criteria-categorised responses depending on the participants' professional backgrounds reveal that all groups cited *Originality* as highly important for GD creativity and *Public-appropriateness* as highly important for PD creativity.

Designers (table 11), are the only group of participants who cited all the six criteria as at least moderately important. Moreover, even if *Originality* and *Aesthetics* are the only ones considered as highly important, they are closely followed by *Public-* and *Brief-appropriateness* (each cited by 50% of participants, so really close to the threshold of highly important criterion). For PD we observed more differences of criteria citing frequency, nevertheless, designers are the group that, compared to others, consensually cited the most of PD criteria.

Table 11: The most cited criteria within the designers group

Criteria of **high importance** (mentioned by >50% of participants)

Criteria of **moderate importance** (mentioned by between 30% and 50% of participants)

Designers - GD		Designers - PD	
Originality	63%	Public-appropriateness	88%
Aesthetics	63%	Aesthetics	69%
Public-appropriateness	50%	Originality	63%
Brief-appropriateness	50%	Brief-appropriateness	38%
Affect	44%	Elaboration	31%
Elaboration	31%		

Within the GD area, art directors shared with designers their preference for *Aesthetics*, while with teachers they shared the preference for *Brief-appropriateness* (both criteria considered as highly important by the mentioned groups). Within the PD area, their

preferences are similar to those shown in the global results. This is the only group that did not share even the moderate consideration for *Brief-appropriateness* within PD.

Table 12: The most cited criteria within the art directors group

Criteria of **high importance** (mentioned by >50% of participants)

Criteria of **moderate importance** (mentioned by between 30% and 50% of participants)

Art directors - GD		Art directors - PD	
Originality	71%	Public-appropriateness	76%
Brief-appropriateness	59%	Originality	65%
Aesthetics	53%	Aesthetics	65%
Public-appropriateness	35%		

Teachers showed the shared interest for lower number of criteria than other groups. For them only *Originality* and *Brief-appropriateness* were highly important for GD creativity, *Aesthetics* coming as moderately important, finally *Public-appropriateness* and *Affect*, contrary to other groups, did not find the teachers' shared interest. Similarly, in the case of PD, teachers only shared a consideration for *Public-appropriateness* as highly important. Contrary to others, they did not share an even moderate consideration for *Aesthetics*. They only did so for *Originality* and *Brief-appropriateness*.

Table 13: The most cited criteria within the teachers group

Criteria of **high importance** (mentioned by >50% of participants)

Criteria of **moderate importance** (mentioned by between 30% and 50% of participants)

Teachers GD		Teachers PD	
Originality	70%	Public-appropriateness	70%
Brief-appropriateness	70%	Originality	40%
Aesthetics	40%	Brief-appropriateness	40%

6.2 Criteria depending on the participants' level of experience in design

When comparing the results depending on the participants' level of experience in design, we note that for GD, *Originality* and *Aesthetics* are shared by more than 50% of each group, while for PG it is the case of *Public-appropriateness* only.

The asserted experts' preferences reflect those of the whole sample (even if their order is slightly modified), with the difference of not including *Public-appropriateness* (table 14).

Table 14: The most cited criteria within the asserted experts group

Criteria of **high importance** (mentioned by >50% of participants)

Criteria of **moderate importance** (mentioned by between 30% and 50% of participants)

Asserted experts - GD		Asserted experts - PD	
Originality	76%	Public-appropriateness	71%
Brief-appropriateness	62%	Originality	67%
Aesthetics	52%	Aesthetics	52%

Intermediary experts cited all the six GD criteria as at least moderately important. For PD it is the case of only one criterion, *Public-appropriateness*, while *Originality* and *Aesthetics* were cited by less percentage of participants (considered as moderately important), compared to other groups.

Table 15: The most cited criteria within the intermediary experts group

Criteria of **high importance** (mentioned by >50% of participants)

Criteria of **moderate importance** (mentioned by between 30% and 50% of participants)

Intermediary experts - GD		Intermediary experts - PD	
Originality	59%	Public-appropriateness	86%
Public-appropriateness	59%	Originality	50%
Aesthetics	55%	Aesthetics	45%
Brief-appropriateness	55%		
Affect	36%		
Elaboration	32%		

Within the GD area, laypeople considered as highly important only *Originality* and *Aesthetics*, while the appropriateness found only the moderated interest. Within the PD area, the choices were similar to those of other participants, but with a different order: laypeople showed more preference for *Aesthetics* than for *Public-appropriateness*.

Table 16: The most cited criteria within the laypeople group

Criteria of **high importance** (mentioned by >50% of participants)

Criteria of **moderate importance** (mentioned by between 30% and 50% of participants)

Laypeople - GD		Laypeople - PD	
Originality	80%	Aesthetics	75%
Aesthetics	55%	Originality	60%
Public-appropriateness	45%	Public-appropriateness	60%
Brief-appropriateness	45%		

7 Discussion

This study allowed us to understand better the nature of (1) detailed evaluative referents and (2) more general criteria, representing the mental representations of design creativity. The findings from this study will be compared with those from other two studies of this thesis. First, we will use criteria cited with the highest frequency by the participants of this study, in order to find out to which degree they influence the overall creativity assessment in the situation of attributing scores to design productions. Second, we will compare evaluative referents expressed as general mental model of design creativity with those spontaneously expressed during the situation of real design assessments.

The first goal of this study was to find out whether there is a set of general criteria that people find important for the design creativity. After analysing our findings, we can propose such a set of the most important criteria, taking into account those cited by more than 50% of our sample. Thus, for GD area, on which we focus in this study, it would be *Originality*, *Aesthetics* and *Brief-appropriateness*. Additionally, we have *Public-appropriateness* criterion was cited by between 30 and 50% of participants, thus, we consider it as a moderately important criterion. Moreover, *Public-appropriateness* seems to be the most important for PD creativity, therefore, for the case of design creativity in general, this criterion should be considered as highly important. These four criteria are tested in the second study of this thesis, with the goal of understanding to which degree participants take them into account while assessing overall creativity of real examples of design.

Two criteria issued from the state of art, were not included in our set of the most important criteria, since they were mentioned by a lower number of our participants: *Affect* and *Elaboration*. It seems that in people's mental representations these criteria do not play the major role. Nonetheless, even if these criteria do not seem to be declared as important for creativity, it does not mean that they are not taken into account during the real situations of design assessment, for which we considered a separated study.

Our second goal was to find out if we can apply the same creativity criteria to different areas of design. More precisely, we compared the areas of graphical and product design. The analysis of responses to our survey revealed that the evaluative referents are similar in both areas, nevertheless, we should keep in mind that each area has its specificities. **The specificity of GD**, when compared to the PD, seems to be in the importance of the conveyed message: the semantic relevance of means and codes for the appropriate presentation of the subject, its understandability, but also its style, related to the designer's personal graphic preferences or the brand identity. **The specificity of PD** is that the message conveyed by the design in this area is rather related to the information about how to use the product. The respect of a biggest amount of constraints and a creation of a certain desire to possess also seem specific to this area. Moreover, we should remember that the physical components and aesthetical means are not the same for these two areas of design and that while in PD one would rather focus on the materials and volumes, in GD more consideration would be given to the colors, 2D shapes, fonts and layout.

The difference between those two areas is also visible in the distribution of the importance assigned to each criterion in terms of significance to creativity. While *Originality* seems to be clearly the most important criterion to GD creativity, for PD creativity it is a case of *Public-appropriateness*, which could be strongly related to the fact that the main goal of

the product is to be attractive towards its usability, while the graphic outcomes do not have the functional dimension and their power of attracting attention is related to the fact that they seem different and outstanding. Thus, we can conclude the productions within the GD area are in the first place expected to be innovative and their creative value will be especially based on this criterion, while in the PD area, more focus will be put on the user-centered criteria.

Moreover, when comparing other criteria that we qualified as highly important, we can find out that, while *Originality* and *Aesthetics* are important for both areas, the appropriateness criteria are considered only partially by each of them: while the PD's creativity depends on *Public-appropriateness* (which was already discussed in the previous paragraph), it seems that creativity in GD is more based on the *Brief-appropriateness*. In GD this criterion is based on two ERs: *Relevance to the subject* and *Respecting the codes of the domain*. Thus, the appropriateness within the GD area seems to be based on the fact that the message of an artifact should convey precisely what it is supposed to convey, using the existing, conventional codes. Since each visual message is full of very specific symbols meanings, it is crucial for the communication source that the message gets received precisely as wanted.

Our third goal was to verify if mental representations of creativity are different depending on the professional viewpoint (background) and on the level of experience in design. We focus especially on results related to the GD area, since they will be developed in further parts of this thesis.

First, we analyzed criteria declared by **participants with regard to their different professional viewpoints**. The most striking difference between the three profiles is that **designers** seem to have the most developed mental model of creativity, since comparing to other groups, they share the most important number of ERs and criteria that they consider important for creative design. While considering only the GD criteria, we can see that *Affect*, *Public-appropriateness* and *Elaboration* are especially more considered by designers, than by other groups. In the PD area this group also shared more criteria than others. This might be linked to the observation made by Bonnardel and Sumner (1996), on designers' professional activities. In lines with these authors, to assess their productions (as well as those of their colleagues), designers have to use multiple perspectives and take into account the viewpoints of different customers. This attitude might also partially explain the particular interest of designers for *Affect* and *Public-appropriateness*, since these criteria are related to focus on the reactions possibly generated by the design.

If we look at the quantity of criteria on which agree **art directors**, we see that it is moderate compared to designers. Art directors' considerations for GD criteria are very similar to those found in general results, which could be interpreted as a validation of this group as being the appropriate representatives of the domain voice. If it's their role to be opinion leader, to select the memes and transmit them to the relevant audience (cf. Harrington, 1999) their choices and preferences should be somehow reflected by the choices and preferences of the domain in general. This is also in lines with Hooker, Nakamura and Csikszentmihalyi (2003): domain gatekeepers should be able to identify the adequate degree of appropriateness and novelty, which means that to identify this adequateness, they should share, or at least have the knowledge about the mental representations of different actors of the field.

Teachers seem to have the strictest vision of creativity: they agreed on the lowest number of criteria in both areas of design. Their approach to creativity is similar to that of art

directors, but comprises less criteria, which confirms that both groups are close to each other and can be put under the common label of domain gate keepers, but with a slightly different focus. In comparison to teachers, art directors' professional role consists more in meeting the audience's values. The results of ERs analysis within the PD area are in lines with Birenbaum (2007), who claimed that teachers are believed to be centered on creativity as an individual process. It seems that teachers developed a system of creativity values that is not only clear (as expected from domain gate keepers), but also restricted to only few shared criteria. The important number of non-shared criteria could be a result of differentiation of the approach to the design developed during the years of teaching experience, during which theory meets the variety of individual pedagogical cases. This could be interpreted in line with Caroff and Besançon (2008), who underlined that some experts could diverge in their opinions about creativity, since their subjective understanding of it was developed differently during the acquisition of their own experience. Nevertheless, we should also remember that the quantity of participants in this group was lower than in other group, which might influence the results.

When **comparing the results of participants with different levels of experience in design**, we can see that the main difference between **laypeople** and the more experienced participants is in their approach to the appropriateness criteria. Laypeople seem to take them into account less than the other two groups and it is visible in both design areas. It might be the effect of their implicit theories about the creativity. Therefore, the professionals of design, following the practical experience that they accumulated, give a high importance to the appropriateness constraints related to the communication of the subject. Laypeople may stay on more superficial level of *Aesthetics* and *Originality*, taking less into account the fact that the original and aesthetic elements have been worked out with the aim of corresponding to specific constraints.

These findings are the most visible within more detailed ER results concerning GD area: we can see that laypeople cited **only** *Innovation* as highly important for creativity. This means that in this group *Aesthetics* and the two appropriateness criteria were built by participants that cited ERs of different nature and were not homogenous in their way of describing the final criteria. Differently from them, asserted experts and intermediary experts agreed on higher number of details concerning GD creativity. We can even observe that asserted experts agreed strongly on a smaller number of criteria, while intermediary experts agreed moderately on a higher number of criteria.

We could advance a hypothesis that while people do not have experience in the creative design domain, they have no many common ERs. With increasing experience the professionals of design find more shared ERs and when they achieved the asserted experts' level, these ERs are reduced to a smaller number, but shared by more individuals. It could be due to the fact that more the experience of people increase in the GD area, more the ERs are shared (thanks to the increasing domain-related knowledge). With time needed to acquire the experience, several ERs lose their importance and the agreement is kept only for those ERs that kept their universality towards different situations that can be encountered during the design professional's career.

Chapter 5. The influence of judges profile on the criteria used in the real situation of the creative design assessment

The goal of this experimentation is to analyse the judges' use of the imposed criteria in a situation of design assessment. It contains two studies, in which we consider two factors of possible variation of creativity assessments: the judges' professional backgrounds and their level of experience

The goal of this experiment was twofold.

First, we aimed to determine if the assessments of graphic design made on the imposed criteria are influenced by the judges' backgrounds and their level of experience.

Then, our goal was to determine the influence of these two factors on the variations in the degree to which the judges are consensual in their assessments on different criteria and on the criteria that influence the most the judges' overall assessments.

1 Pilot study in the field of product design

Before conducting this experimentation in the field of graphic design, we conducted a similar study in the field of product design (Wojtczuk & Bonnardel, 2012). It allowed us to present some methodological issues as well as some results obtained within the previous study in order to show the development of the experimental design for the final study that we present here.

Our goal was to obtain productions created on the basis of the same brief. The best way to acquire a relatively large number of productions by people with the same level of experience and on the same topic, was to cooperate with design students. For this study in the field of product design, we asked them to design a new computer mouse. We gave them one hour to come up with the 3-D mock-ups. It seemed to us that mock-ups already represent a design proposition that can be assessed by judges. Nevertheless we came to the conclusion that mock-ups are hardly comparable with the finished designs, since they contain neither the feasibility details, nor for the finishing work. Furthermore, the judges assessing the mock-ups complained about the lack of possibility of trying out and touching the real versions of presented products. They argued that the functionality is the most important criterion for assessing product design and since mock-ups were not really functional, it was hard to assess them according to this criterion. This seems to be an important constraint for studying the assessment of product design, since even if we had had more time, we could not obtain the finished products because of the expensive production process. Thus, we decided to conduct a research with the same goal, but in another field of design.

In the case of graphic design, the productions obtained by design students, even if not presented in their real, final format, are more comparable to the final products. The production process of graphic designs requires no other tools than a computer with graphical software. To assess graphic productions, the judges do not need the final print of the graphic design work, they can formulate an opinion assessing a digital version or with a small A4 printed version.

The method used for the previous study was as follows: We asked 20 judges with four different types of professional backgrounds (professional designers, design teachers,

retailers and users), to assess 18 product designs (the mock-ups of computer mice). We asked them to assign scores from 1 to 7 for each assessed design. More specifically, first they assigned scores with regard to four specific criteria that were chosen basing on other research in the field (*Aesthetics, Originality, Functionality and Marketability*) and then, the score for overall assessment.

The aim of this study was to determine if there is an impact of judges' backgrounds on final design assessments. ANOVA measures revealed that there was no significant effect of this factor on the overall assessment $F(1, 16) = 1.18, p > .05$. However, when we explored the detailed assessments based on specific criteria, we observed a significant effect of the background on the functionality score $F(1, 16) = 3.70, p < .05$. The retailers seemed to be stricter in their functionality assessments than the professional designers, who had a tendency to give high scores on this criterion.

Moreover, we analysed partial correlations between all the overall assessment scores and the specific criterion scores for every group of judges to find out which criteria seemed to be the most relevant for overall assessments, depending on the judges' backgrounds (table 17). Since each of five judges within each group made 18 assessments, we obtained 90 assessments within each group of judges. We uncovered similarities between retailers and users: both these groups seemed to base their overall assessment mainly on Functionality ($pr = .60, p < .001$ for retailers and $pr = .56, p < .001$ for users), and on Marketability ($pr = .58, p < .001$ for retailers and $pr = .50, p < .001$ for users). By contrast, the assessments of the designers and design teachers seemed to depend more on *Aesthetics* ($pr = .52, p < .001$ for designers and $pr = .47, p < .001$ for teachers). The *Originality* criterion effect on the overall assessment was generally the weakest of all four specific criteria. Nevertheless, this criterion seemed to have some importance for designers ($pr = .48, p < .001$).

Table 17: The distribution of partial correlations between the specific criteria scores and the overall scores, depending on the judges' professional background. The strongest partial correlations for each group of judges are marked in blue

	Designers	Design teachers	Retailers	Users
Functionality	.44	.42	.60	.56
Aesthetics	.52	.47	.41	.46
Originality	.48	.34	-.02	.47
Marketability	.43	.37	.58	.50

The analysis of results confirmed the initial hypothesis that judges base their assessments on different criteria, according to their background. The judges could be divided into two main groups: those who were involved in creative activities (designers and design teachers) and those who focused on practicalities (retailers and users). These results allow us to extend Norman's model (1988). Indeed, more specific analysis of results for each participants profile led us to make the following observations:

It seems that **users'** mental representations of objects are based more on their Functionality. This is not surprising, as they are supposed to use the objects that are being assessed. Moreover, the marketing trends and attractiveness seem also to influence their product design assessments.

By contrast, **designers** seem more interested in the *Aesthetics* and *Originality* of the objects. This shows that when assessing a design product, they are particularly interested in the product's shape and the innovative idea.

Retailers, as well as concentrating on the Marketability aspects, clearly have a good understanding of the users' perspective: their assessments mirrored the users' concern with Functionality, although they were far stricter.

Finally we observed that the correlations between different criteria were weakest with the **design teachers'** scores. This can be attributed to a desire to give precise and detailed judgements, paying particular attention to each criterion. Furthermore, in the course of their work, teachers are often asked to justify their judgments when assessing students' work. We also noted that they seemed to share the designers' perspective concerning the importance of *Aesthetics* in their overall assessments.

1.1 General goals of the experimentation

The experiment presented in this chapter is inspired by the previously described pilot experiment, with some changes in the experimental design.

The present study was conducted within the field of graphic design, which had some consequences on the profiles of recruited judges and on the criteria used to assess the designs, since, as we could see in the study presented in Chapter 4, the criteria to assess the design creativity and their order of importance differ depending on the field of design (graphic design vs product design).

Moreover, the present study aims to study differences in assessment of design creativity. Because of this, to measure the overall assessments we did not use the "overall score", like in the pilot study, but the "overall creativity score". Our goal was to examine links between the specific criteria and the creativity of a design, while in the pilot study participants giving their overall assessments probably expressed their personal preferences, with no special concern about the creativity.

Moreover, we aimed to analyse not only the influence of judges' different backgrounds on creativity assessments, but also to know if these assessments are also influenced by the judges' level of experience in the field of design. Thus, this experiment contains two separate analyses of two factors:

The first study examines the influence of the judges' professional backgrounds on the creativity assessments of graphic design.

The second study investigates the influence of the judges' level of professional experience in the design field on the creativity assessments of graphic design.

Both studies are based on the same experimental procedure, in which judges are asked to assess 21 posters, rating them on **overall creativity** and on four specific criteria relevant to visual communication design. The criteria for assessing posters' creativity were chosen following the results of the study 1 of this thesis. We took into account only the criteria that were quoted by more than 50% of participants: *Originality*, *Aesthetics*; *Brief-*

appropriateness and *Public-appropriateness*. Therefore, besides the overall creativity score we asked judges to assign scores with the regard to these four specific criteria.

2 Study 1: The influence of judges' backgrounds on the graphic design assessment

In this study we aim to understand the differences of creativity assessments of graphic design (in our case, the posters), that depend on the backgrounds of judges.

Following to the findings described in the pilot study (Wojtczuk & Bonnardel, 2012), the judges' overall assessments of product design are influenced by specific criteria of different nature, and these differences are influenced by the judges' backgrounds. These findings develop Norman's model (1988), which illustrates users' and designers' differences of perception of the same product, by showing that these differences also concern other actors in the design field, especially those, playing the role of the domain gatekeepers (design teachers and design retailers). We estimate that the these differences of perception can be also observed within the graphic design assessments of creativity, performed by judges with different backgrounds, all potentially involved in the process of graphic design: **designers** involved in creation; **art directors** involved in approving and managing the design projects; **design teachers** specialised in design training; and the **audience** that represents the potential target.

Moreover, as we could observe from results presented in Chapter 4 of this thesis, judges from different backgrounds declared some criteria to be more appropriate than others for assessing the creativity of a graphic design. All of them proposed *Originality* as important for creativity.

We can expect that judges with the corresponding profiles would show the same preferences when assigning scores on the creativity during the assessments of real designs. The aim of this study is to determine, taking into account the variations of numerical scores, the nature of these preferences.

2.1 Hypotheses

Our **general hypotheses** are the following:

Our first general hypothesis is that the **backgrounds** have an **impact** on the way that judges assess graphic designs creativity.

Moreover, we expect to find the differences depending on the judges' backgrounds within two aspects:

Our second general hypothesis is that the variations of the **inter-judge agreements** on the scores assigned for specific criteria are influenced by the judges' backgrounds.

Our third general hypothesis is that the specific **criteria that contribute to the overall creativity assessments** are different, depending on the judges' backgrounds.

Our **specific hypotheses** concerning both, the degree of inter-judge agreement and the criteria the most contributing to the overall creativity assessments, are the following:

H1. According to Amabile (1983), judges do not have to agree on the definition of the creativity if they agree on what is creative. Nevertheless, within the design field, the creativity criterion could be perceived as too general, since many authors prefer to assess it using lists of more specific criteria. If it is more efficient to make assessments on the specific criteria rather than on the overall creativity, **we expect the judges' assessments on specific criteria to be consensual to a higher degree than those on overall creativity.**

H2. In line with Norman's differentiation between users' and designers' mental representations (Norman, 1988), as well as the results of the pilot study (Wojtczuk & Bonnardel, 2012), we expect the **criteria that guide the most the creativity assessment of art directors to be shared with the audience's ones** (since professionally they are supposed to know the audience's preferences in the field of design), while **design teachers will share those of designers** (since they shape the designers' mental representations of creative design).

H3. In line with Hooker, Nakamura and Csikszentmihalyi (2003), **domain gatekeepers** should possess enough domain knowledge to identify the required degree of appropriateness and of novelty, which should have as consequence the **higher level of inter-judge agreement on all the proposed criteria** for art directors and teachers than for other groups.

H4. Concerning the designers' results, we expect their creativity assessments to be **based especially on Originality** criterion – since their perception of creativity might be influenced by their constant research for new ideas (Bonnardel, 2000, 2006) and **on Public-appropriateness**, because of their professional habit of taking into account the customers' and users' viewpoints (Bonnardel & Summer, 1996). Moreover, according to the Lera's conclusion (1981) on the fact that designers attribute the importance to the project's constraints in heterogeneous way, we expect within them to obtain a **low inter-judge agreement on Brief-appropriateness.**

H5. Concerning **the audience**, in line with the theory of implicit theories (Chan & Chan, 1999), we expect their **creativity assessments** to be based on their random, personal system of beliefs rather than on shared, explicit theories and in consequence **less based on specific criteria** and **with lower inter-judge agreement.**

2.2 Participants

We selected 20 men and women whose backgrounds allowed a balanced distribution over the four following groups:

- 5 professional designers, with 9.4 years of professional experience as designer on average;
- 5 design teachers, with 12.4 years of teaching experience on average;
- 5 art directors (who make decisions about style and overall visual appearance, and give directives to designers), with 10.8 years of professional experience on average;
- 5 participants with no experience in the design domain, representing a potential audience of the posters (balanced age and sex).

2.3 Material

We collected 21 graphic posters designed by students in the same year of study at a graphic design and visual communication school. All the posters had to be on the same topic, the goal being to inform the potential audience about a specific event concerning ecological packaging. We chose a population of students not only for ease of recruitment (it would be difficult to find enough professionals available at the same time) but also to avoid differences in expertise (on the basis of their education level).

The students were given 2 months to come up with their final results and the project was a part of the obligatory syllabus. These conditions allowed us to obtain the productions resulting from a long-time creative process, which comprised several constraints, like requirements expressed by the teacher, deadline, possibility of interacting with different sources of inspiration.

All the posters were digitalised and downloaded on a specific experimental website.

2.4 Procedure

The assessments were made online. Each judge was given an individual ID and password to access the experimental website. The 21 posters were displayed in random order. We asked the judges to rate all the posters on all the criteria, by giving scores of 1-7.

First, the judges rated the posters on overall creativity. Then, all the posters were shown to them again, so that they could rate each of them on the four specific criteria.

Short definitions were provided for each criterion, to avoid differences in interpretation:

Overall creativity of a product is related to its originality as well as to respect of the task or constraints;

Brief-appropriateness depends on both the respect of constraints, the relevance of the transmitted message and the way in which the theme is presented;

Aesthetics refers to the aesthetic finishing, the composition, the aesthetic pleasure felt while looking at the product;

Originality corresponds to the innovative aspect of the product;

Public-appropriateness is related to the marketing efficiency, the capacity of the product to attract the targeted audience.

We collected 462 sets of scores (each of the 22 judges assessed all 21 posters), every set containing one score for overall creativity and four scores for the specific criteria.

2.5 Results

We carried out three types of analyses in order to know the differences between judges' assessments, depending on their backgrounds.

2.5.1 The general impact of judges' backgrounds on the assessments of the design creativity

To find out if there is an impact of judges' backgrounds on final design assessments, we first conducted an **ANOVA**. Its measures revealed that there is a highly significant effect of the judges' backgrounds on the overall creativity assessment $F(1, 16) = 16.93, p < .001$.

It also revealed highly significant effects of the judges' backgrounds on the four criteria: *Aesthetics* $F(1, 16) = 17.5, p < .001$; *Public-appropriateness* $F(1, 16) = 10.57, p < .001$; *Brief appropriateness* $F(1, 16) = 10.81, p < .001$; *Originality* $F(1, 16) = 23.83, p < .001$.

With these results, we can conclude that depending on their background, judges attributed the scores on the overall and specific creativity criteria in different ways.

2.5.2 The impact of judges' backgrounds on the degree of the inter-judge agreement

In order to verify the consensual character of assessments, we used **Cronbach's alpha**, commonly used to estimate the internal consistency for sample of participants. Amabile justifies the use of Cronbach's alpha coefficient by the fact that it is a static indicator generally used to valid the internal reliability of questionnaires and psychometric tests, while the estimated degree of agreement can be understood as reliability, since it estimate the agreement of the same set of judges on other products.

Table 18 shows the results for Cronbach's alpha depending on their backgrounds. It can vary between 0 and 1. Most authors consider that used this measure for their research in creativity consider that the α higher than .7 presents a good consensual agreement, while the α between .5 and .7 presents a weak degree of agreement (Amabile, 1996, Hennessey, 2004, 2009; Kaufman et al, 2004 et 2008, Baer et McKool, 2009).

When comparing the values of Cronbach's alpha from table 18, we can observe that no group of judges reached the $\alpha > .7$ for the overall creativity assessment, while each group is highly consensual on one specific criterion: Art directors agree mostly on the *Aesthetics* criterion, designers on the *Originality* criterion and both teachers and audience agree on the *Public-appropriateness* criterion.

We can observe that the judges' assessments are not highly consensual for the majority of criteria. Mostly, the weak agreement is only achieved, especially for teachers and audience, since both groups achieve $\alpha > .7$ for one criterion and $.5 < \alpha < .7$ for three criteria. Designers seem to be the less consensual group, since they agreement is low for the majority of criteria.

Table 18: Distribution of inter-judge agreement (expressed in Cronbach's alpha) within posters assessments based on 5 criteria depending on judges' backgrounds.

	$\alpha > .7$ good consensual agreement			
	$.5 < \alpha < .7$ weak consensual agreement			
	Art Directors	Designers	Teachers	Audience
Creativity	0,59	0,42	0,62	0,68
Brief-appropriateness	0,09	0,49	0,46	0,64
Originality	0,63	0,75	0,52	0,57
Public-appropriat.	0,45	0,58	0,70	0,73
Aesthetics	0,74	0,19	0,69	0,49

2.5.3 Differences within the criteria that contributed to overall creativity assessments, depending on the judges' backgrounds

We aimed to compare the specific criteria that contributed the most to the overall creativity assessments expressed by the judges with different backgrounds. Thus, we took into account all the assessments (a single assessment is the one made for a single poster by a single judge, on one overall and four specific creativity criteria) and compared which of the four specific criteria scores covaried the most with the score assigned for the overall creativity. We made these analyses using **linear regression**, separately for each group of judges with different backgrounds.

The results of these analyses revealed that, depending on the judges' backgrounds, the four criteria had different effects on the creativity assessments for each group of judges. The table 19 illustrates the distribution of the most significant effects.

For **designers**, we found highly significant effects of *Originality*, $b = .35$, $SD = .08$, $\beta = .37$, $p < .001$, and *Public-appropriateness*, $b = .33$, $SD = .08$, $\beta = .36$, $p < .001$.

For **design teachers**, we found highly significant effects of *Originality*, $b = .35$, $SD = .06$, $\beta = .35$, $p < .001$, and *Public-appropriateness*, $b = .27$, $SD = .06$, $\beta = .31$, $p < .001$. For *Brief-appropriateness* we found a significant effect, $b = .21$, $SD = .07$, $\beta = .2$, $p < .01$.

For **art directors**, we found a highly significant effects of *Brief-appropriateness*, $b = .52$, $SD = .07$, $\beta = .55$, $p < .001$. The significant effects was found for *Aesthetics*, $b = .2$, $SD = .07$, $\beta = .2$, $p < .01$, and *Public-appropriateness*, $b = .2$, $SD = .07$, $\beta = .21$, $p < .01$.

For the **audience**, we only found a tendency effect of *Brief-appropriateness*, $b = .26$, $SD = .11$, $\beta = .3$, $p < .05$.

Table 19: The distribution of significant effects of the specific criteria scores on the overall creativity score, depending on the judges' professional background.

*** highly significant effect

** significant effect

* tendency effect

	Designers	Design teachers	Art directors	Audience
Brief-appropriateness		**	***	*
Aesthetics			**	
Originality	***	***		
Public-appropriateness	***	***	**	

2.6 Discussion

The general influence of judges' professional backgrounds on the creativity assessments in graphic design.

ANOVA results revealed significant differences between the creativity scores awarded by the judges with different backgrounds. This is in line with the **first general hypothesis**. These findings show that the “Background” factor is playing an important role in the variations of design creativity assessment. Other results allow us to find more details about these dissimilarities.

We tested the **second general hypothesis** about the differences within the inter-judge agreement on the scores attributed by judges with different backgrounds. Cronbach's alpha analyses are in line with this hypothesis, since the criteria on which the judges' assessments are consensual differ depending on the judges' backgrounds.

Linear regression analyses by group revealed that depending on their backgrounds, for each group of judges there is a different criterion that contributes the most to the overall creativity assessment. As a consequence, the **third general hypothesis** is validated.

Moreover, according to **hypothesis 1**, judges' assessments on specific criteria should be consensual to a greater degree than those on overall creativity. The results partially confirm the first hypothesis, since we find some specific criteria with a higher degree of agreement than for the overall assessment, but this rule does not concern all of the specific criteria.

If we take into account the findings on these variations of inter-judge agreements, it seems that agreements on specific criteria are different depending on the background of judges who assigned the scores. This observation suggests that judges with the same background develop some kind of objectiveness, but one that is restrained only to the specific criteria, which results in more consensual assessments on these criteria. This could be a confirmation of the general hypothesis of this thesis: people can develop different approaches to design creativity, with a special regard to specific criteria, as consequence of the specific mental representations created by during their specific professional activity.

The results verify **hypothesis 2**, since designers seem to follow the same principles in assessing creativity that design teachers: for both of them, *Originality* and *Audience-appropriateness* guided the overall creativity assessment. Moreover, art directors seem to share their way of assessing creativity with audience, since for both of those groups *Brief-appropriateness* contributed most to the overall creativity assessment.

These findings confirm the results obtained by the pilot study (Wojtczuk & Bonnardel, 2012) as well as our hypothesis about Norman's model (1988) according to which there is a distinction between a “creative” profile and those with a “user” profile. This model seems to be valuable not only for usability studies, but also for the assessment of creativity in graphic design. Thus, designers and teachers seemed to focus on the creative search, while the audience and art directors seemed to focus on the informative function of the poster described in the brief, which need to be easy to understand by the targeted audience.

Hypothesis 3 was not validated: the two profiles that we chose to represent domain gatekeepers (art directors and design teachers) and should share common domain knowledge, did not agree on scores given on all the creativity criteria to the higher level than other groups. Even if teachers' results contain a relatively high amount of weak agreements, they do not seem to share this tendency with art directors – they show surprisingly more common points with audience. We presumed that gatekeepers are those who are

recognized in their domain and through accumulating a sufficient amount of design knowledge, have a common perception about the creativity of design products and its specific criteria. Nevertheless, the preferences of this group were as specialized as for other judges: teachers obtained a high agreement only on their *audience-appropriateness* judgments and art directors only on *Aesthetic* ones.

Nevertheless, these results could be interpreted as supporting our major hypothesis on the influence of judges' backgrounds on their mental representations about creativity. We could suppose that the gatekeepers' role of those who accept or reject creative works, shapes their mental representations of creativity in a way that makes them agree in their judgments on specific points, rather than expand their agreement on all the criteria that concern creativity, making of them "the most objective judges of creativity". The fact that their agreement on overall creativity is also quite low could only confirm that hypothesis about the specialisation of judgment as a consequence of a particular background.

We should take into account the fact that even if the gatekeepers from our study did not agree on a high number of criteria, their overall creativity assessments were significantly guided by the most of these criteria. Both, teachers' and art directors' judgments were influenced by *Brief-* and *Audience-appropriateness* and a third criterion, which was: *Originality* for teachers and *Aesthetics* for art directors. The two remaining groups' assessments were influenced only by one or two criteria. Therefore, we could infer that thanks to their domain knowledge and their important positions in the world of design, gatekeepers have more tendencies to use multiple criteria for creativity assessments, compared with judges with other backgrounds, even if the scores attributed on them are not always consensual. The teachers' approach might be shaped by their practice of using multi-criteria forms (Reis & Renzulli, 1991). Art directors might be trained to this multi-criteria approach by the responsibilities of their decisive positions which comprise the interactions with both: executive designers and clients.

We should notice the importance that art directors attach to *Aesthetics*: not only does it influence their overall creativity assessments; furthermore, the character of judgments made on this criterion seems to be highly consensual. This could be the effect of their high experience with different types of aesthetic trends in design, and the particularity of their work, in which they have to make choices between various propositions made by designers and building the aesthetic style of their clients' identity (typography, colours, layout, etc.).

We have validated **Hypothesis 4**. As expected, designers' assessments were mainly based on *Originality* and *Audience-appropriateness*. We supposed that the creative process of research for new ideas (Bonnardel, 2000, 2006) and taking their users' viewpoint on their own design (Bonnardel & Sumner, 1996) could influence the way in which the designers approach not only their own outcomes, but also those of their peers.

For the second part of this hypothesis, an interesting point was found. We observed that, as expected, designers presented a low degree of agreement when assessing designs on *Brief-appropriateness*. Nevertheless, they were not the only group with this approach. A low degree of agreement is found on the *Brief-appropriateness* within all the groups of design professionals. This is surprising, since we might expect that it would be easiest to agree on the criterion that requires a comparison of the artefact with the brief, which was a common reference, accessible for each judge during the whole experiment. We could expect that the *Originality* criterion would be more difficult to judge in a consensual way, since the possible references can vary and depend on the judges' individual past experiences. This dissonance could be in line with Lera's findings (1981) assuming that while designers manage the brief constraints, they can attribute the importance to the brief elements in

differring ways, depending on their personal values or their understanding of the customers' values. It might be possible that this behaviour can be generalized to judges with other backgrounds than just designers.

The **Hypothesis 5** was validated only in part. When analysing the **audience's** results, we can see that none of the specific criteria strongly influence their overall creativity assessments (since only one tendency effect was found). This could reveal the lack of a deeper involvement in the understanding of design creativity, a characteristic of implicit theories (Chan & Chan, 1999). The lack of detailed knowledge could be the reason for which the audience participants associate the concept of creativity to specific criteria to a lesser degree than judges who have a background connected with professional design.

Nevertheless, the audience's relatively high (compared with other groups) inter-judge agreement on the majority of criteria is not in line with our hypothesis. This results are rather in line with findings suggesting that the non-experts' creativity assessments can be sufficiently consensual (Dollinger, Urban, & James, 2004; Hickey, 2001; Kaufman et al., 2005). When attributing scores on criteria, audience participants use their "lay-references", which are probably less influenced by domain-specific knowledge. It seems that the common references of this group are less a source of disagreement than the specific references of the professionals. Moreover, we should note that the audience shows the highest agreement on *Audience-appropriateness*, which is not very surprising and somehow shows that the relevance of this criterion (since participants representing the design target agrees on what is attractive for them).

3 Study 2: The influence of judges' level of experience in the field of design on the assessment of graphic design

In this study, we aimed to find out the nature of the differences within creativity assessments of judges with different levels of professional experience in design.

Most authors defend the hypothesis that the nonexperts' inter-judge agreement is lower than experts' agreement (Kaufman et al., 2008; Lee, Lee, & Young, 2005; Plucker, Holden, & Neustadter, 2008). Nevertheless, some authors obtained opposite results, where laypeople obtained higher inter-judge agreement than experts (Dollinger, Urban, & James, 2004; Hickey, 2001; Kaufman et al., 2005). In both cases, we can see that experts assess creativity differently from laypeople, which was proven by the low correlations obtained between the ratings of these groups of judges (Hickey, 2001; Kaufman et al., 2008). In this study, we do not only aim to confirm the results about experts' or nonexperts' higher levels of agreement, but also to find out about the nature of these differences, expressed by criteria that contribute the most to creativity assessments.

These differences were already discussed in Chapter 4 of this thesis. We obtained results that led us to conclude that participants with different levels of experience in design attach different levels of importance to different criteria, when considering which of them are important to assessing creativity. We can expect that these differences will be also visible if we examine, in a real situation of design assessment, which of specific criteria most impact the scores, assigned by judges with different levels of experience in design, for overall creativity.

Moreover, our participants are divided into three groups: **asserted experts** with more than 10 years of experience in the field of design, **intermediary-experts**, with between 5 and 10 years of experience in this field and **laypeople** with no experience in the field of design. Thus, our study allows not only to analyse differences between the highly experienced judges and those who have no experience at all, but also to see if there are any real differences between those who exceeded the 10 years limit of expertise in the field of design and those who have less of experience than 10 years, but are not laypeople. The 10 years limit was determined by several authors (Bloom, 1985; Gardener, 1993; Hayes, 1989) and this study aims to find out how it influences creativity assessments (to what degree the experts and the laypeople differ in their way of attributing scores for design creativity).

3.1 Hypotheses

Our **general hypotheses** are the following:

Our first general hypothesis is that the **level of professional experience** in design has an **impact** on the way that judges assess graphic designs.

Moreover, we expect to find differences depending on the judges' backgrounds within two aspects:

Our second general hypothesis is that the variations of the **inter-judge agreements** on the scores assigned for specific criteria are influenced by the judges' level of experience.

Our third general hypothesis is that the specific **criteria that contribute to the overall creativity assessments** are different, depending on the judges' level of experience.

Moreover, as we could observe from results in the study 1 of this thesis, judges from some groups may prefer some criteria more than others. The aim of this study is to determine, taking into account the variations of numerical scores, the nature of these preferences.

Our **specific hypotheses** concerning the degree of inter-judge agreement and the specific criteria contributing to the overall creativity assessments are as follows:

H1. Similarly to the first study, we expect the judges' assessments on specific criteria to be more consensual than those on overall creativity.

H2. Following the findings by Kaufman et al. (2008), Lee, Lee and Young (2005) or Plucker, Holden and Neustadter (2008), the inter-judge agreement is lower for nonexpert judges. Thus, we expect that the asserted experts and intermediary experts will show a higher degree of agreement for all the assessment criteria and for the overall creativity assessment than laypeople will.

H3. In line with authors who argue that only people who have more than 10 years of experience in a domain can be considered as experts, we expect to find differences within the degree of inter-judge agreement between asserted experts and intermediary experts.

3.2 Participants

We selected 21 men and women with different levels of experience in graphic design. We obtained a balanced number of participants in each of the following three groups:

- 7 asserted experts in design – with more than 10 years of experience in design, with 19.3 years of professional experience in design on average;
- 7 intermediary experts in design – with between 5 and 10 years of experience in design, with 6.6 years of professional experience in design field on average;
- 7 laypeople– with no experience in design (balanced age and sex).

3.3 Material

We presented to participants the 21 graphic posters designed by students in graphic design and visual communication (the same posters as in the previous study). The posters were informing the potential audience about an upcoming event concerning ecological packaging. They were all presented on a specific experimental website.

3.4 Procedure

As in the previous study, assessments were made online, each judge connecting with an individual ID and password and assigning to all posters scores of 1-7, on one overall creativity criterion and 4 specific criteria (*Brief-appropriateness*, *Aesthetics*, *Originality* and *Public-appropriateness*). 441 sets of scores were collected (each of the 21 judges assessed all 21 posters). Every set contained one score for overall creativity and four scores for the specific criteria.

3.5 Results

We proceeded to three types of analyses in order to identify the differences within the judges' assessments, depending on their level of experience.

3.5.1 The general impact of judges' level of experience within the design field on the assessments of the design creativity

We conducted an ANOVA in order to determine if differences of distribution of mean scores assigned by the different groups of judges for the different criteria are statistically significant or not. Results revealed that there is a highly significant effect of the judges' level of professional experience in design on the overall assessment of creativity $F(1, 18) = 28.21, p < .001$.

The results also revealed highly significant effects of the judges' level of professional experience in design on the four criteria: *Aesthetics* $F(1, 18) = 21.47, p < .001$; *Public appropriateness* $F(1, 18) = 7.39, p < .001$; *Brief-appropriateness* $F(1, 18) = 9.08, p < .001$; *Originality* $F(1, 18) = 16.11, p < .001$.

3.5.2 The impact of judges' level of experience within design field on the degree of the inter-judge agreement

In order to verify the consensual character of assessments, we used **Cronbach's alpha**. Table 20 shows the results depending on the participants' level of experience.

Table 20: Distribution of inter-judge agreement (expressed with Cronbach's alpha) within poster assessments based on 5 criteria depending on judges' level of experience in design

	$\alpha > .7$ good consensual agreement		
	$.5 < \alpha < .7$ weak consensual agreement		
	Asserted experts	Intermediary experts	Laypeople
Creativity	0,64	0,52	0.65
Brief-appropriateness	0,54	0,2	0.59
Originality	0,74	0,77	0.67
Public-appropriateness	0,69	0,68	0.74
Aesthetics	0,49	0,64	0.55

As we can see, no group of judges reached the $\alpha > .7$ mark for the overall creativity assessment, while each group is highly consensual on one specific criterion: asserted experts and intermediary experts agree the most on the *Originality* criterion, while laypeople agree mostly on the *Public-appropriateness* criterion.

For other criteria (except *Brief-appropriateness* for intermediary experts and *Aesthetics* for asserted experts), the agreement between judges is weak.

3.5.3 Differences in the criteria that contributed to overall creativity assessments, depending on the judges' level of experience

Using **linear regressions**, we compared which of the four specific criteria contributed the most to overall creativity scores for each group of participants, depending on their level of professional experience in design field.

Table 21 illustrates the distribution of the most significant effects found with linear regressions analyses for each group of judges.

For **asserted experts**, we found highly significant effects of *Originality*, $b = .38$, $SD = .06$, $\beta = .38$, $p < .001$, and *Public-appropriateness*, $b = .20$, $SD = .06$, $\beta = .23$, $p < .001$. For *Brief-appropriateness* we found a significant effect, $b = .22$, $SD = .07$, $\beta = .23$, $p < .01$.

For **intermediary experts**, we found highly significant effects of *Originality*, $b = .25$, $SD = .59$, $\beta = .28$, $p < .001$, *Public-appropriateness*, $b = .38$, $SD = .07$, $\beta = .40$, $p < .001$ and *Brief-appropriateness*, $b = .23$, $SD = .56$, $\beta = .26$, $p < .001$.

For **lay-persons**, we found a significant effects of *Brief-appropriateness*, $b = .28$, $SD = .09$, $\beta = .3$, $p < .01$.

Table 21: The distribution of significant effects of the specific criteria scores on the overall creativity score, depending on the judges' level of professional experience in design.

*** highly significant effect

** significant effect

* tendency effect

	Asserted experts	Intermediary experts	Laypeople
Brief-appropriateness	**	***	**
Aesthetics			
Originality	***	***	
Public-appropriateness	***	***	

3.6 Discussion

General influence of judges' level of professional experience on the graphic design creativity assessments

ANOVA results revealed significant differences between the creativity scores awarded by judges with different levels of professional experience in design, which is in accordance with the **first general hypothesis**. These findings confirm that the “level of experience” is the source of variation within assessments of design creativity.

The **second general hypothesis** involved differences in inter-judge agreement on the scores attributed by judges with different levels of experience. The Cronbach's alpha analyses are only partially in line with this hypothesis: the asserted experts' and the intermediary experts' assessments are consensual for the same criterion – *Originality* – while laypeople agree on *audience-appropriateness*. Thus, only the laypeople can be considered as agreeing on different criteria, compared to the whole sample. Therefore, we can conclude that there is a difference between people having at least some experience in design and people having no experience in this field, if we consider the nature of specific criteria on which they achieve a high level of inter-judge agreement.

Linear regression analyses by group revealed that the four criteria had differing effects on creativity assessments for each group of judges. Thus, the **third general hypothesis** is validated.

According to **hypothesis 1**, the judges' assessments on specific criteria should be consensual to a higher degree than those on overall creativity. Our results partially confirm the first hypothesis, since we found some specific criteria with a higher degree of agreement than for the overall assessment, nevertheless, the judges do not agree highly on all the specific criteria.

To verify **hypothesis 2** – asserted experts' and intermediary experts' agreement coefficients were expected to be higher for all the criteria than the agreement of laypeople

– we compared Cronbach’s alphas for the three groups of participants with different levels of experience in design. The second hypothesis cannot be validated, since the laypeople’s coefficients of consensual agreement are not lower than those obtained in other groups.

Nonetheless, when comparing the results of asserted experts and intermediary experts with those of laypeople, we note that even if the inter-judge agreements are not very different in their values, they are of a different nature. Participants who have some professional experience in design agree on the *Originality* criterion, whereas those who have no experience at all in this field agree on *audience-appropriateness*.

Nevertheless, we observe that **asserted experts’** and **intermediary experts’** highest agreements are of the same nature: **hypothesis 3**, about the differences within the assessments of asserted experts and intermediary experts, cannot be validated. Both groups agree on the *Originality* criterion, and thus their degree of inter-judge agreement does not seem to vary depending on whether they reached the limit of 10 years in their professional experience or not. Therefore, their results are clearly different from those of laypeople, who agree mostly on *Audience-appropriateness*. It is possible that it is enough to have just some professional experience in design, to accumulate the knowledge necessary for having similar references about already existing designs and to establish a shared limit between what is original and what is not.

Chapter 6 – Criteria used spontaneously in real situation of creative design assessment

In this study, we analyse more precisely the criteria involved in the assessment of creative design by studying the criteria used in real situations of assessment.

Since this study offers a wide range of qualitative data, we have structured it in three separate stages. During the first stage, we present the results of data categorization, in order to better understand the criteria used by judges in the situation of assessing creative designs. During the second stage, we focus on the criteria used spontaneously by judges, in different types of assessments of creative design outcomes. During the third stage, we focus on the influence that the judges' backgrounds and their levels of experience in design exert on their assessments.

Contrary to the first study presented in this thesis, this study does not focus on the criteria understood as declared principles and mental representations of the creativity in design, but on the criteria deduced from evaluative referents, used in practice, when judges refer to their own concepts of creativity to assess real-world designs.

Moreover, we wish to study the assessment criteria with the consideration for both the visceral and the reflective levels of design (Norman, 2004), but in our case we applied these levels to the perception of design creativity. We considered that in the case of graphic design, we should explore only two of Norman's three levels (see state of art). Hence, we assumed that the behavioural level is only linked to the product design, because of its physical function and usability, while in the case of graphic or print design, it is difficult to observe the participants' interactions with the creative outcomes. We took into account 1) the visceral level – to observe the participants' first reaction to the presented designs and, in line with the goal of our study, its level of creativity, and 2) the reflective design – to observe their in-depth analysis, with regard to the judges' knowledge and values about design creativity, applied on the real examples. Therefore, the design of this study was tailored to allow the observation of the evolution of the assessment process, taking into account as well the positive as the negative opinions resulting from each use of evaluative referents.

Data collection for this study was based on semi-structured interviews. The participants were assessing the provided examples of graphical designs and had to express their opinions about them, using their own understanding of what is creativity in design, without receiving suggestions to use any specific criteria. They were asked to think aloud when analysing the designs' creativity and to make choices of outcomes that they considered to be the most creative. By structuring the interviews into four stages, we placed our participants in four different situations of assessment: (1) choices based on the visceral assessment of creative design, (2) reflective assessment of creative design (detailed analyses of good and bad points perceived within each design), (3) choices based on reflective assessment of design creativity, (4) choices based on personal preferences concerning design.

The analyses were conducted for each stage of interviews in order to understand which criteria are decisive in different situations of creative design assessment. First, we analysed the interviews using a qualitative method in order to find and label the criteria used for assessments. After that, we verified the frequency of their occurrences within the entire sample of participants as well as within the groups representing different professional viewpoints and levels of expertise in design.

Several goals are set for this study:

- 1 To observe which criteria are the most frequently used in a real situation of creative design assessment.
- 2 To compare the criteria used for choices based on the visceral and reflective creativity assessments and those based on personal preferences in the field of graphic design.
- 3 To analyse differences and commonalities between the criteria used by judges with different professional viewpoints and with different levels of expertise in design.

1 Methodology

Our aim was to encourage our participants to think aloud and express as much evaluative referents as possible to assess the creative designs. The semi-structured interview method seems to be the most suited to this objective. It has some main advantages for our study:

- Participants express their opinions in real time, which allows us to identify the elements of their process: the participants' first reactions, the changes occurring during the entire process and the explanations of some statements.
- Participants can use the vocabulary that comes spontaneously to their minds and describe their thoughts without having to choose between specific, imposed criteria. Moreover, the experimenter can ask about more explanations if some of the used terms seem ambiguous.
- The collected data reflect the participants' spontaneous choices and preferences, since they are based only on the criteria they thought of, not the imposed ones.
- The context of assessment is taken into account, which gives the ecological validity to our data.

Nonetheless, this method presents also several weaknesses:

- There is always a part of experimenter's personal understanding in the way in which qualitative data are analysed, even if the categorization process comprises several experimenters.
- The verbalizations contain a large amount of information that is lost during the categorisation process.
- The presence of the experimenter can induce a change in the way in which the participants naturally behave and express their opinions.
- The fact of verbalizing one's thoughts can influence the natural way of thinking.

Even taking into account these weaknesses, conducting the interviews nevertheless seems to be the best way to gather information. This includes the entire process of assessment comprising the verbalized evaluative referents actually used by participants.

1.1 Data analysis

The data collected during the interviews were organized within four groups, corresponding to the four stages of our study. We recorded the verbalisations of our participants and

wrote them down in order to have the written texts as material for qualitative analyses. We analysed each group of data by adapting the ground theory method (see chapter 3). We proceeded as following for each stage of our study:

- We identified **evaluative referents** by collecting terms used by participants to assess outcomes. Since different terms were identified as expressing the same evaluative referents, they were classified within the same group.
- We clustered the evaluative referents with similar contents, in order to find the common concepts. The groups are categorised and labelled as operative **criteria**.
- These criteria can be compared with findings from other stages of this study, with other studies presented in this thesis and with findings resulting from other research.

Using this methodology, we collected a large number of elements concerning the processes occurring in real situations of creative design assessment. (1) Our data included the criteria spontaneously used in the context of creative design assessment; (2) we could find out which of these criteria were used most frequently; (3) we compared the quantity and the nature of criteria mentioned by participants with different professional viewpoints and levels of expertise; (4) we made a comparison between the criteria of choices based on visceral and reflective assessments of creativity with those based on personal preferences.

2 Participants

2.1 Recruitment

We conducted interviews with 23 participants. 16 of them were professionals of design who agreed to be available for an interview lasting at least 30 minutes. They were mostly recruited in Marseille and Aix en Provence, but some of them were from Nîmes and from Paris. In addition we interviewed also 7 laypeople in order to compare the assessment criteria used by experts with those used by participants having no experience in the design field.

2.2 Establishing specialization profiles

We chose participants with the same profiles as for the first study. Our aim was to obtain four groups:

- Designers
- Art directors
- Design teachers
- Laypeople

As in the first study, we had to define the decisive criteria to know to which group we should assign participants with multiple professional activities. As in the first study, we decided that being an art director is a decisive criterion, since it changes the viewpoint on the domain. Thus, we separated participants who were designers only from those who were also art directors. We did the same for design teachers. Consequently, even if a participant was not only a teacher, but also a designer, we qualified him or her as a teacher.

2.3 Establishing experience profiles

We also wished to observe within our sample some differences related to the participants' levels of professional experience in design. As in the previous study, we choose as the limit between more and less experienced participants a duration of 10 years of professional experience in the design domain. On these bases, we created the following groups:

- Participants with more than 10 years of professional experience in the design domain, that we will name **asserted experts**.
- Participants with no more than 10 years of professional experience in the design domain, that we will name **intermediary-experts**.
- Participants with no professional experience in design at all, that we will name **laypeople**.

2.4 Final sample

Our participants were recruited with regard to their professional profile and the level of their professional experience. We interviewed:

- 6 **designers** who are not teachers or art directors (4 males and 2 females, average number of years of experience = 9.7. 2 of them are asserted experts and 4 of them are intermediary experts)
- 4 **art directors** who are not teachers (4 males, average number of years of professional experience = 15.5. 2 of them are asserted experts and 2 are intermediary experts)
- 6 **design teachers** (6 males, average number of years of professional experience = 16.3. 4 of them are asserted experts and 2 are intermediary experts)
- 7 **laypeople** (4 males and 3 females, in different age groups)

To obtain data about the degree of experience in design, within the same population, we created 3 groups:

- 8 **asserted experts** (8 males, average number of years of professional experience = 14.3. The group includes 2 designers, 2 art directors and 6 design teachers)
- 8 **intermediary experts** (6 males, 2 females, average number of years of professional experience = 7. The group includes 6 designers, 2 art directors and 2 design teachers)
- 7 **laypeople** (4 males and 3 females, in different age groups).

3 Material

In order to provide the graphic designs to assess, we selected 8 graphic posters designed by students in the same year of study at a graphic design and visual communication school. All the posters were on the same theme, concerning an event about ecological packaging. The 8 posters were selected between the 21 works from the study described in chapter 5. Based on the results of the previous study, we choose 3 posters with the highest scores attributed for creativity, 3 with the lowest ones and 2 that were judged as highly creative within only several groups of judges, while other groups gave them the lowest creativity scores.

All designs were colour-printed in an A4 format and presented to participants in random order. Each participant was also given a printed brief that students used to create their designs.

4 Procedure

4.1 Interviews

The interviews were conducted in different places, depending on the availability of designers (mainly their offices and their homes). All of them were recorded and written down.

Judges were told to express their opinions about the designs, to think aloud expressing each thought coming to their minds concerning the assessed design, even if their opinions could appear incoherent. In order to avoid too much kindness or an overly pedagogical approach, they were told that none of designers who produced the assessed artefacts would be informed about what was said concerning his or her work.

The interview was conducted in four stages:

1. After a short presentation of the study and an introduction to the brief, the participant was asked to look through the 8 posters and to choose those that he or she thought the most creative. There was no limit for number of chosen designs, so that the judges could feel free to make spontaneous choices.

At the end of this stage, the judges were asked to explain the reasons for which they have made their choices.

2. The judges were shown the same posters than in the previous stage, but this time they saw them one by one (still in random order) taking their time to analyse the good and bad points of each design and to express their opinions. For each poster, judges were asked two questions:

- What do you think about the creativity of this poster?
- What would you change in this poster to make it better?

The first question was an open one to encourage judges to express their opinions without giving them any specific directions. The second one was asked in order to formulate the critics in a clear way and to obtain complementary data.

3. We asked judges to choose again the posters that they considered the most creative, but taking into account the previous analysis. In order to obtain the responses with more ecological value (as if the judges were making the real life choices), we asked them which of the posters they would recommend to be really used for being exhibited in the city, to inform about the event.
4. Finally, judges were asked to choose the posters that they personally preferred, without thinking of creativity or of the brief, for example, which poster would they choose to decorate their apartment.

4.2 Analyses

Since our data are the verbatim collected during semi-structured interviews, we were able to proceed mostly only to the qualitative analyses. Therefore, this study has an exploratory

character; we describe our findings and the possible tendencies that we could deduce from the analyses of our data in three stages:

During the first stage, we analysed the verbatim to identify the evaluative referents (ERs) used by judges to assess the design creativity. For this, we used our adapted version of grounded theory methodology: we identified codes, categorized them in groups that expressed the ERs, themselves being grouped under even more general labels of criteria. Finally we could compare these findings with those from previous research.

During the second stage, we analysed the number of occurrences for each ER (and for criteria in which they are included), in order to understand and compare their distribution during different types of assessments. Of course, we cannot assume that the ERs that are used most often are the most important ones, but we can study the granularity of creativity assessments and the changes of the judges' focus during different types of assessments.

During the third stage, we analysed differences in the use of those ERs and criteria, between judges with different backgrounds and different levels of experience. Still, without being able to make conclusions on the degree to which these results are significant, we explore our findings in order to find interesting tendencies, which could eventually give inspiration for further studies.

5 First stage of the study: categorisation results

At this stage of the study we did not take into account differences between our participants. We considered the total verbalizations from all the interviews in order to identify ERs used to assess the creativity of presented graphic designs.

5.1 Method of analyses

We wish to keep the process of data analysis in line with the grounded theory, which means that our categories are first constructed with regards to field data, and next only compared with already existing theories.

5.2 First step: identifying codes

All the interviews were recorded and analysed. We treated the four stages separately, but using same procedures.

While analysing an interview, we identified the parts of judges' verbatim reports in which they expressed their judgments (positive or negative judgments) about the graphic designs that were presented to them.

5.3 Second step: categorisation and ER finding

Next, we carried out the categorisation process. A group of three persons was involved in this procedure: two researchers in the field of creativity and one professional designer. They assigned the verbatim to the general categories and indicated the name for each category. The common categories were built using verbatim from all four stages. Each category was labelled and used for the further quantitative analysis as an Evaluative Referent (ER) of creative design assessment.

We wished to calculate a number of occurrences for each ER. By *occurrence* we mean an application of a specific ER within a specific judgment of a particular poster. That means that a single judge can mention the same ER several times, while assessing different posters. Moreover the judge can use the same ER for assessing the same poster, but at different stages of our interview. Each use of this ER is considered as a separate occurrence. Nevertheless, if a judge mentioned an ER several times while assessing the same poster during the same stage of interview, we considered it as a single occurrence – in order to discard situations in which more talkative judges used synonyms while meaning the same ER.

We approved a double occurrence of the same ER within a single judgment only during the second stage of the interviews, while judges were commenting good and bad points of posters (while during other stages they were mostly explaining their choices, thus, commenting mostly their positive point). Thus, the same ER can result in a positive or negative statement. We took these statements' values into account and treated them as separated ER occurrences.

5.4 Third step: criteria finding

We compared the list of obtained ERs with criteria used for assessing creativity that exist in other research. In order to obtain data that were more general and more comparable with existing research, we grouped ERs matching already-known criteria.

After this analysis, we obtained a list of criteria and by summing up the number of occurrence of the corresponding ERs, we obtained the number of occurrences of each criterion. Moreover, we had access to the following information:

- The number of occurrences of each criterion within each group of judges;
- The number of positive and negative statements resulting from the use of each criterion during the reflective assessment;
- The frequency of use of each criterion in design choices resulting from different types of creativity assessments (visceral *vs* reflective) and from personal preference.

5.5 Identification of ERs

During the categorisation process 15 categories of ERs were defined. We present their descriptions and some examples of related quotations, extracted from interviews, illustrating the positive and negative statements that resulted from the application of each ER. The quotations revealed the context of the ER use and provided us with additional details, compared to study #1, which operated only with ERs built from single terms proposed by participants as evaluative referents, without any assessment context.

Concept development – comprises the visible sign of idea elaboration, of an in-depth reflection about the theme, the proof that the author carried out some intellectual work to find a way of presenting the subject in an interesting way.

Positive statements:

Il y a une vraie idée (...) On sent la volonté que l'affiche exprime le salon de l'écologie, de ne pas passer à côté du message. (M - 2)

Cette idée-là a été poussée jusqu'au bout, il n'y a rien à enlever rien à rajouter (Jf – 1)

Au moins c'est une idée et elle est assumée (Jf – 5)

En termes de créativité, il a essayé de conceptualiser la notion de la boîte. Il va faire un clin d'œil, pour lui la boîte c'est l'œuf (F - 8)

Negative statements:

Il n'y a pas vraiment d'idée, c'est difficile à décrypter. On ne sait pas vraiment de quoi il s'agit. (Ma -3)

L'idée du code barre ce n'est pas assez poussée, Ça manque de travail, c'est superficiel (Tp - 3)

Il y a une conceptualisation, mais ce n'est pas assez assumé, La piste autour de la boîte pourrait être amusante, mais là, elle est timide, on a envie que ça aille un peu plus loin (F-3)

Quelqu'un a pompé un visuel qu'il a vu sur Internet, donc pour moi ça n'a pas d'intérêt. On est avec un élève qui veut être à la mode, mais qui n'en retire rien. Un copier-coller d'un blogue ne fait pas partie d'une démarche conceptuelle. (F - 6)

Innovation - as in previous studies, innovation refers to the original and uncommon side of design. We found two specific viewpoints related to the graphic design innovation:

- **Idea innovation** – refers to the used symbols that should be out of common, surprising, something that we are not used to, avoiding banality, common codes, something that everybody could think of, easy interpretations.

Positive statements:

On pense à la nature, mais pas stéréotypée (...) on se détache du carton, de ce qu'ont fait les autres (O – 5)

Il y a des symboliques assez intéressantes (Fr - 2)

Negative statements:

Je trouve ça très scolaire (...) montrer l'arbre pour montrer l'écologie, ce n'est pas très original... ces signes manquent d'originalité (M – 3)

C'est green et en plus c'est une box, on est au premier degré (S - 4)

Les éléments sont là, mais si je paie quelqu'un pour ce travail je dirais non, car je peux aller aussi bien dans une bibliothèque d'images, prendre un arbre, mettre greenbox, un papillon, car j'ai déjà vu ça quelque part et puis je mets de la pelouse derrière (Re - 3)

- **Graphic innovation** – for inventiveness in graphic choices. Similarly to the idea innovation, it focuses on avoiding banality and common codes on the graphic level (e.g. choices of colors, fonts, etc.)

Positive statements:

On montre le carton comme matière, support de typo, c'est intéressant (Tp – 3)

Les jeux sur la perspective, c'est intéressant (J -2)

J'aime bien la composition qui change un peu des affiches trop centrées (Tp – 5)

La police crayonnée, c'est une typo alternative (S – 6)

Negative statements:

Je vois comment c'est fait, il a utilisé des motifs tout faits de Photoshop (Tp - 2)

Ce n'est pas parce qu'on parle d'écologie qu'il faut du vert (N - 5)

Les arabesques on les a vues et revues, il ne faut plus faire ça (Ma – 2)

Risk-taking appreciation – associated to the use of uncommon strategies, where the courage to make radical choices is appreciated, but with concern about the appropriateness.

Positive statements:

On touche à quelque chose de fragile, compliqué, tabou (M – 7)

Là, il y a quelqu'un qui essaie d'être un peu plus moderne, qui s'est dit : tiens, je vais prendre des risques, je vais mettre des corps (F – 7)

J'aime bien le parti pris graphique, c'est très fort (Au - 8)

Negative statements:

J'attends autre chose de quelqu'un qui est frais, pétillant, avec une liberté de pensée, car il n'est pas dans un contexte professionnel (M - 3)

On est sur la douce provocation mais il faudrait avoir vraiment le courage de ses actes, ce qui n'est pas le cas ici (F – 7)

C'est limite de mauvais gout (A - 7)

Aesthetics appreciation – used to describe general aesthetic impressions, without analysis of graphical details.

Positive statements:

C'est quelque chose qui a une tenue de point de vue graphique (T - 3)

C'est sophistiqué, le graphisme est bien traité (A - 3)

Negative statements:

Elle est créative mais pas élégante, pas très belle, le graphisme est en péril (E - 2)

Graphiquement c'est nul, je donne ça à ma fille, elle fait pareil (M - 7)

Design elements – are mentioned during the detailed aesthetic analysis, with a consideration for the different elements used in graphic design. We detailed the elements that were frequently quoted: **Fonts, Texture, Contrast, Colour, Background, Illustration quality, Logotype**. In their positive or negative statements, judges simply commented the design elements as good or bad choices.

Positive statements:

Il y a un gros travail de typo, ce n'est pas facile, c'est bien de l'avoir tenté (Tp – 3)

On voit un joli jeu de texture (A – 3)

Je ne sais pas comment ça a été fait, mais il y a un vrai travail sur le visuel (Ma – 5)

Une vrai greenbox qui est logo-typée, on peut le mettre sur le site web, sur les murs, flyers etc. (Au-5)

Negative statements:

A mon sens, le traitement typographique est complètement maladroit (Fab - 5)

La texture est un peu ratée, ça fait bois du parquet (Tp – 5)

Ça manque de contraste : tout est dans les tons marron (Au – 2)

Pourquoi du vert et du marron ? Coloristiquement parlant c'est un désastre (N - 8)

La photo de base est mauvaise : on part d'une mauvaise matière première (A – 6)

Quality of execution – comprises a technically sound work with a professional effect, with a visible effort, a professional approach and compliance with rules and techniques used for treatment of graphical elements.

Positive statements:

C'est très bien, car il faut toujours que ce soit qualitatif pour le respect de la marque (Jf – 4)

Ce qui sauve l'affiche, c'est qu'elle est propre et bien faite, c'est bien réalisé (J - 4)

Negative statements:

Le graphisme peut être très simple, mais là ça fait amateur (E - 8)

C'est vraiment un brouillon. Il y a une faute d'orthographe, il faut faire un effort, on dirait qu'il n'y a rien de maîtrisé (S – 7)

Bonne idée mais pas une bonne réalisation, la réalisation casse tout (J – 5)

Layout – the way in which the whole work is composed: the placement and proportions between images, logos and texts, an appropriate hierarchy, with no overload and not too much empty space.

Positive statements:

La personne n'a pas eu peur du blanc, ça apporte de la fraîcheur, de l'espace, on respire (Au-4)

J'aime bien la composition qui change un peu des affiches trop centrées (Au - 5)

Negative statements:

Tout le blanc autour, c'est soit pas assez soit trop, il faudrait caler différemment (N - 2)

On n'arrive pas à poser son œil, aucun élément ne cale notre regard sur la proposition visuelle (N – 6)

Au niveau de la mise en page il s'est trompé : le positionnement de la bouteille sur le côté gauche, donc sur le côté pauvre de l'affiche, alors qu'on voudrait balayer l'affiche de haut en bas. Il faudrait peut-être inverser de côté : mettre les infos sur la gauche et mettre en avant le visuel (N - 5)

Harmony – the way of choosing different elements that gives to the final work an impression of unity. For this ER, we found only negative statements.

Negative statements:

L'idée de la typo est intéressante, mais en quoi les papillons sont en relation avec la forme ? (N - 3)

Il n'y a pas d'intégration du texte avec le visuel. On aurait pu imaginer que le texte était gravé au fer chaud sur le bois – on assume jusqu'au bout le produit en bois, mais lui il n'a pas fait l'association des deux (M - 5)

Les personnages ne sont pas à la même échelle, la femme est plus présente, le corps d'homme est parfait et le corps de femme l'est moins... Il n'y a pas de cohésion sur les visuels qui sont choisis (H - 7)

Simplicity – graphical concept that aims for sobriety and avoiding overload, minimalism, purity and clear composition.

Positive statements:

Une pureté sympathique, sans le fond qui fait perdre le contenu de l'image, l'image est mise en valeur (R - 4)

C'est propre, il n'y a pas beaucoup de choses, mais ça suffit (L - 5)

Negative statements:

C'est trop gourmand. Il y a trop de choses. On veut montrer tout à la fois (M - 2)

Le graphisme est trop compliqué : trop de superpositions (J - 2)

Graphic style – this ER groups the statements that judges made about used graphic means of expression (for example: contemporary, childish, hippie, institutional, actual, old-school).

Positive statements:

On est sur une représentation très classique et haut de gamme, à cause de la typographie ancienne, les papillons sur le côté, visuellement c'est très vieillissant. On peut jouer là-dessus (N - 3)

Elle est sympa, car utilise le vocabulaire de collage, un peu malhabile, papier canson, découpé... un peu enfantin, qui va bien (S - 8)

Negative statements:

C'est propre, côté catalogue, il y a un maniérisme qui n'apporte à rien (M - 4)

Graphiquement c'est un registre un peu enfantin, ça me plaît pas trop (Fr - 8)

Comprehension of the message – includes the comprehension of symbols that should be easily understood by the targeted audience and the readability of the most important information.

Positive statements:

Très claire, je comprends qu'il s'agit d'un salon sur l'écologie, parce que j'ai ce volume-là très lisible, on parle du végétal, donc c'est accessible... (A - 4)

On voit bien que c'est une bouteille, en plus elle est en bois, donc on a le côté écologique. Une bouteille est un objet qu'on reconnaît tous (L - 5)

Negative statements:

Est-ce qu'on fait un rapprochement entre la fragilité de l'écosystème et celle de l'œuf... ? Le cheminement est compliqué au premier abord dans un abris-bus (O - 8)

On ne sait pas pourquoi on a affaire à un bernard l'hermite, il n'est pas associé à l'écologie. Ce serait un panda qui sort de la boîte... (N - 1)

Normalement sur le code barre, on a des chiffres ; Si on prend un stéréotype, il faut le coller vraiment au stéréotype visuel (N - 2)

La lecture est un peu compliquée, je ne l'ai comprise qu'après avoir lu 2 fois : Pourquoi un œuf, pourquoi fragile, pourquoi l'évènement est dans le blanc d'œuf – quand on se pose trop de question, ce n'est pas bon (Jf - 8)

Relevance to the subject – the degree to which the design fulfils the goals stated in the brief (especially, relevance with the subject), by the use of appropriate metaphors and the presence of all mandatory themes (ecology, packaging, event).

Positive statements:

C'est celle qu'aurait pu choisir le client. On voit les éléments, on comprend qu'il s'agit d'écologie (H - 3)

Ca suit le cahier des charges pour créer de nouveaux gestes quotidiens (A - 4)

C'est pas l'idée du siècle, mais ça fait son boulot (Au - 5)

Cette affiche-là, je la trouve très pertinente par rapport au visuel utilisé, qui est à la fois très rigolo, touchant, qui parle assez bien d'écoconception ou d'éco-emballage (Jf - 1)

Negative statements:

Il manque le caractère informatif, il faut respecter le cahier des charges, car ce n'est pas une œuvre d'art (B - 4)

Et qu'est-ce que me dit que je suis dans un salon d'emballage ? Il n'y a rien. Juste le texte (M - 3)

Est-ce que ça suffit à suggérer les enjeux de cette journée ? On peut être agacé, car ça ne répond pas aux contraintes de la communication (R - 7)

C'est hors sujet, ça ne touche pas à la problématique de l'emballage (E - 7)

Appeal – salient characteristics that catch eye and attract attention of the public, the visual appeal, something that would evoke questioning or create another positive reaction.

Positive statements:

J'ai presque envie de voir un dessin animé avec ce petit animal, Ça m'évoque plein d'histoires (Jf - 1)

Ça joue sur le code couleur : tonique, on est attiré par ce que l'on voit, on se demande ce que ce homard fait là (O - 1)

L'idée est bonne, elle fonctionne et peut être très marquante – on sait tout de suite de quoi ça parle, car on a tous utilisé et jeté des bouteilles en plastique (N - 5)

Negative statements:

Une affiche qui passera inaperçue, donc elle manque sa cible, les gens ne la regardent pas, elle n'interpelle pas car elle n'est pas assez créative (E - 6)

Evoking emotions – includes cases in which judges mentioned any emotional reactions, like appreciation, humour, enthusiasm or joy.

Positive statements:

Elle fait sourire, ce qui fait rire, on le retient mieux (Tp - 1)

J'aime le fait que l'affiche soit dynamique, cette impression du mouvement avec les feuilles. Ça donne l'idée de voir de la pêche autour de la thématique, de vouloir changer les choses (A - 5)

Un apport du vivant donne à l'emballage un côté optimiste. L'emballage n'a normalement pas d'importance, alors que là, ça devient la maison d'un animal (R - 1)

Negative statements:

C'est trop froid pour un sujet sympa, il faudrait plus d'histoire. J'aurais rajouté un truc un peu plus décalé, un diabolin... une petite maison... (Jf - 4)

Je ne l'aime pas. Elle n'est pas pleine d'enthousiasme (J - 5)

Le marron c'est triste, alors que l'éco-emballage devrait être quelque chose de positif pour encourager les gens à venir (J - 2)

References – this ER is slightly different from the others and will not be comprised within the criteria. We wish just to take note of the fact that judges often expressed analogies and connotations with other cultural memes, which were not indicated in the brief.

Positive statements:

J'aime bien, ça me renvoie à Dali (M – 1)

Il y a des références comme green-peace, Woodstock, assez babacool... (H – 8)

Negative statements:

Ça me fait penser aux journées autour de la mer, pas à l'écologie (Au – 1)

Des motifs de prolongements des codes-barres, qui nous renvoient vers les pires cauchemars (T – 2)

5.6 Observations made during verbatim analyses

Originality. It seems that for the *Originality* ERs, the accent is put on the fact that the author of the design went through a creative process: chose the idea which is not stereotyped, found the uncommon graphical means and was consequent in their use. While the *Innovation* ER is more about the use of symbols and graphical elements that are not the stereotypes or first ideas that come to everybody's mind, the *Concept development* is about the intellectual effort made to defend these original choices and to awaken the interest in the observer's mind. To be original, the author should take some risks. Nevertheless while some judges appreciate it, others can misunderstand the final outcome, dislike it or even ridicule it. It seems that in this place passes the border between two main parts of creativity: its innovative - and at the same time, suitable - character.

For example, one of the authors presented human bodies with as only clothes underwear made of grass. Some judges found it an interesting idea, since it was very different from other posters and could evoke different interpretations. Nevertheless many judges criticized the poster severely, telling that this shows only bad taste and takes profit of easy marketing tips that catch the audience's attention by showing half-naked human bodies.

Aesthetics. The Aesthetics appreciation is the only ER that treats aesthetics from a general viewpoint. The judges expressed here their feelings about the global appreciation of graphic quality, without entering into the details and professional terms. The six following ERs are more focused on specific aspects of posters' aesthetics. This aspect of creativity assessments seems to have the highest granularity. Some judges could spend long minutes on analyzing different elements of the poster, taking into account their originality and appropriateness to the aesthetic trends and styles, to the subject and to other elements of the same poster. Within the same poster some elements could be judged positively, some others negatively.

Appropriateness of metaphors. A closer look to the last two ERs reveals that the same element of the graphic design can be judged as appropriate or not by different judges, depending on their values. For example one of the posters represents a hermit crab, which has a specific characteristic of adopting the shells left by other sea animals, as its own refuge. On the poster, the hermit crab uses a packaging instead of shell, which is perceived by some judges as a very good, creative idea, that illustrates the idea of well reused packaging, the use of living animal inspiring good emotions (e.g. *Usually, we do not care*

about the packaging and here it gained an importance, because it became a house of an animal), while other judges perceived it as a symbol that gives the unappropriated connotations to the seafood or maritime environment, with no reference to ecology or packaging (e.g. *We do not understand why there is a crab on this poster. If we talk about ecology, we should rather use something as stereotyped as a panda*).

We propose here a list of used metaphors that were supposed to represent ecological packaging and the degree of their acceptance from different judges:

Hermit crab	11 positives/ 13 negatives
Barcode	11 positives/ 3 negatives
Tree on a box	13 positives/ 4 negatives
Box made of leaves	8 positives/ 2 negatives
Wooden bottle	16 positives/ 4 negatives
Transparent box on cardboard	9 positives/ 9 negatives
People in underwear made of grass	2 positives/ 18 negatives
Broken egg	6 positives/ 15 negatives

Analogies and emotions. The fact that we created a separated category of ERs labelled *References*, is the consequence of our statements made during the analyses of verbatim utterances: it revealed how much the evaluation of graphical design is accompanied by analogical thinking. The analogies encountered during the assessments can be positives or negatives to the final assessments, which depends on the previously described relevance to the subject and on the conveyed emotions. For example, one of the posters represents a bar code that is transformed into flowers. Most of judges perceived in it a positive message (e.g. *there is a will to represent consumption, which can grow ecological*), but some were driven to the much more dark connotations (e.g. *this makes us think about our worst nightmares!*).

5.7 Criteria-finding

Next step was to compare these results with those presented in other research: those made within the previous chapters of this thesis, as well as those made by other authors.

As previously mentioned, research on creativity pointed out the importance of such criteria as *Originality*, *Public-* and *Brief-appropriateness*, *Aesthetics*, *Affect* and *Elaboration*. In chapter 4, we respected these criteria to group evaluative elements proposed by participants as important to assess creativity within these more general criteria. We qualified them as mental representations about creativity. In the present study, we aimed to proceed similarly and to group ERs mentioned by judges in real situation of design assessment

Nevertheless, after the analysis of ERs mentioned in the present study, we realized that some of them are different from mental representations of creativity, presented in the first study of this thesis. We added *Risk-taking appreciation*, which could be linked to the person-creativity ERs, rejected by us in the first study as not connected with product-creativity and we decided to group it within an *Originality* criterion with *Innovation* and *Concept development*.

Also other ERs connected with mental representations of creativity were not revealed by the analysis of ERs mentioned within real-life assessments. Thus, our current study does not comprise *Relevance to the subject*, *Respecting codes of the domain* and *Ergonomics and user-appropriateness*. Moreover, after the comparison made between *Relevance to the subject* and *Comprehension of the message* in the current study, we stated their high similarity and assumed that both of them should be placed in the same category. Therefore, in the current study, we reduced *Brief-appropriateness* and *Public-appropriateness* criteria to the single criterion *Appropriateness*.

Our analyses showed that what we called in previous chapters *Public-appropriateness* is more mentioned in terms of *Appeal* and *Emotions conveyed*. Therefore, we named the remaining criterion *Affect*.

Finally, we found out that *Quality of execution* that was previously connected with *Elaboration*, could be also connected with *Aesthetics*. The same goes for *Style* that we interpreted in the first study of this thesis as belonging to *Originality*, nevertheless, the analysis of context in which this ER was used, showed that it is closer to *Aesthetics*. Therefore, this criterion is the one that contains the highest number of ERs.

Table 22. presents the categorization of ERs for the current study.

Table 22: Categorization of ERs with linking to the creative design criteria from literature

Criteria	Originality	Appropriateness	Affect	Aesthetics
ERs	<ul style="list-style-type: none"> - Concept development - Innovation - Risk-taking appreciation 	<ul style="list-style-type: none"> - Relevance to the subject. - Comprehension of the message 	<ul style="list-style-type: none"> - Appeal - Emotions conveyed 	<ul style="list-style-type: none"> - Design elements - Harmony - Simplicity - Aesthetics appreciation - Layout - Graphic style - Quality of execution

5.8 Conclusion

The results of our qualitative analyses are rich and can give a real insight into the judges' thinking about creativity and their focus on different criteria. The goal of analysing verbalisations is to obtain the context for the ERs and for the criteria that we use in more quantitative studies. Therefore, we can find complementary meanings for ERs and criteria presented in the previous studies of this thesis, like those concerning the judges' mental representations of creativity (see chapter 4) or their lack of agreement on several criteria (see chapter 5).

Categorizing the judges' ERs led us to reconsider our approach of the *Appropriateness* criterion and its division into *Brief-appropriateness* and *Public-appropriateness* used in the previous parts of this thesis. It seemed more convenient to consider that an appropriate way to present the subject is the one that will be understood by the audience, nevertheless the values to be represented are those described in the brief – these are the constraints that should be taken into account by the designer and presented using the most appropriate (comprehensible) analogies. Thus, the *Comprehension of the subject* is more connected to *Brief-*, than to *Public-appropriateness*. *Public-appropriateness* would instead focus on the

degree to which it seems to be attractive to the audience, rather from the emotional than semantic viewpoint.

High granularity of ERs connected with *Aesthetics* could drive us to a question about the scores attributed by judges to this criterion in other studies. What is meant by statements about good or bad aesthetics? Are the judges considering the same elements? Which elements are more important than others to decide about the final score? While talking about design details, like fonts or layout, are they looking for the arguments to defend their first impression, or are they collecting data to give their final opinion at the end? It seems that even though the *Aesthetics* criterion plays an important role in the judgment of creativity in graphic design, there is a space for another study in this area, especially while we realize that each element of the design can be assessed for its creativity (it should be at the same time original and adapted, but adapted not only to the context, but also to the actual rules and trend in the graphic design).

We observed in the creativity assessments the presence of both, analogical thinking and analysis of constraints (like those described in the brief or anticipated within the ability of the audience to understand the message). This is in line with the creative process described by Bonnardel (2000, 2006) in her model of *Analogy and Constraint Management*. It seems to confirm that the assessment is not only a part of the creative process, as described in models by Wallas (1927), Gelb (1996) or Amabile (1996), but also, that both processes (creativity and assessment) have some similar characteristics. In particular, analogical thinking and taking into account the constraints. From our qualitative analysis of verbatim utterances, we stated that *Originality* and *Appropriateness* are often treated as two different sides of the same element: each analogy might be judged as too original but inappropriate (the symbol used might be too far from the stereotype and induce the misunderstanding of the whole message) or too appropriate, and in consequence not original (the used symbol is too close to the stereotype, or was used by others on multiple occasions).

The differences in judgments on *Appropriateness* (while different judges focused on the same element of the design, but some of them assumed it as being suited to the brief and others as not being suited) can give us more insight about why we obtained such a low inter-judge agreement on the *Brief-appropriation* criterion in the second study of this thesis. It reminds the reader of a statement by Lera (1981) about design professionals' assessments that, if they are not precisely briefed by their customers', rely on their own values. It seems that in the case of graphical design, the use of symbols might drive the judge to different connotations, depending on their mental representations of the domain represented by the poster and their understanding of the values represented by the subject.

We also observed that to assess creativity, judges used analogical thinking, but the same element can drive them to different analogies. It seems that the problem of symbol interpretation influences judgments made with regards to *Appropriateness*, but also within those made on *Originality*. The same symbol can drive to different connotations and involve different emotions, which can influence the global reaction to the design as positive or negative. We have no means, in this study, to analyse the process that leads to these different interpretations of graphical design, but we can state that the use of symbols is an important source of variation within the creativity assessments of graphic designs.

6 Second stage of the study: criteria used by judges for different types of assessments.

At this stage we identified which criteria were cited most frequently to assess the creativity of graphic designs. Moreover, we aimed to find out which criteria were mostly used during different types of creativity assessments: the visceral and the reflective criteria, as well as the assessments based on the judges' personal preferences.

The study is designed so that we could be able to observe the distribution of the ERs used across different types of assessment. For the first type of assessment, the visceral type, we asked participants to choose designs that they considered to be the most creative “after the first approach” and to explain their choices without entering into too much detail. For the second, the reflective one, we asked them to look again at each design and explain to which degree they find it creative and for which reasons. These two types of assessment are based on Norman's theory of different levels of design (2004), following which, on the visceral level, people tend to be influenced by their emotions and the first aesthetic impression, while on the reflective level, they take into account values that they find important for the domain in which they are making their assessments. In this part of the study we did not ask the participants which designs they liked, which would be more connected to Norman's theory, but we asked which design they found most creative, in line with their personal understanding of creativity. Therefore, we expected to understand on which ERs visceral and reflective assessments of designs' creativity are based. We wished to have a “control situation” and to compare the criteria used to assess creativity and those used for assessments based on personal preferences toward the design. For this, we asked participants to make and to explain their choices that reflect which designs they personally liked the most.

The results obtained by analysing participant verbalizations show the number of occurrences of each ER and should instead be compared on the basis of ERs, not on the basis of criteria. In the following parts of this chapter we present the distribution of both ERs and criteria, but we should remember that in this study the criteria represent the categories of summed-up occurrences of ERs, not the single occurrences of each criterion per judge and per poster. Some criteria include many more ERs than others, which increases the frequency of their occurrence, which makes statistical comparison between them impossible, but gives us more detailed information about the nature of the judges' assessments. Even if the results obtained in this way cannot be compared with those obtained in the previous parts of this thesis, since they are not of the same nature, the analysis of the following results, allows us to have an overview of several interesting tendencies.

6.1 The distribution of criteria used by all judges during different types of assessments

Our goal was to understand **which criteria are the most frequently used during the assessment process** of the creative design, with no differentiation between different levels of judgment. For this reason, we compared the percentages of criteria occurrences, used by judges during different types of assessments (visceral, reflective and based on personal preferences).

First we are going to examine the global results, obtained with the totality of criteria cited during the entire interview. Next, we will compare the results obtained within the specific types of assessment occurring during the interview.

6.1.1 Global results

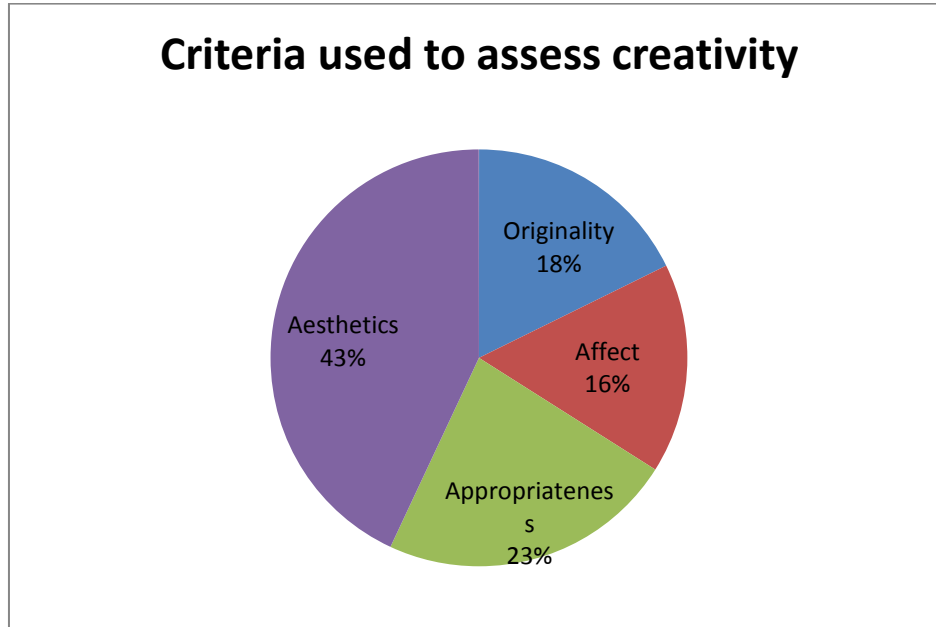


Figure 6: Percentages of occurrences of criteria used by judges during the entire interview.

Originality	296
Concept development	84
Innovation	198
Idea innovation	162
Graphical innovation	36
Risk-taking appreciation	14
Aesthetics	719
Aesthetics appreciation	49
Design elements	325
Fonts	101
Texture	34
Contrast	18
Colour	53
Background	42
Quality of illustration	61
Logo	16
Quality of execution	97
Layout	158
Harmony	18
Simplicity	47
Style	25
Appropriateness	384
Relevance to the subject	145
Compreh. of the message	239
Affect	272
Appeal	92
Emotions conveyed	180

Table 23: Criteria and ERs used by judges during the entire interview

First, we compared the percentages of occurrences of all the criteria, used by all judges during all the stages of interviews, applied for all the posters. To understand the meaning of each criterion, we analysed which ERs it comprised.

Figure 6 shows the distribution of criteria used by the whole sample. It seems that the criterion mentioned most frequently by participants is *Aesthetics*. If we look closer to the ERs (Table 23), we can see that, when talking about this criterion, judges seem to attach much importance to the details of *Design elements* (especially to fonts, to quality of illustrations and colors) to the *Layout* and to the general *Quality of execution*.

The Appropriateness criterion is much less mentioned than Aesthetics, nevertheless, we should taking into account that it comprises Comprehension of message, which is the second most frequently mentioned ER (after *Design elements*).

The *Originality* criterion mostly involves *Idea innovation* and *Affect* mostly depends on *Emotions conveyed*.

Next step was to compare the distribution of criteria on which were based different types of assessments. To analyse this, we compared the numbers of occurrences of criteria mentioned during each type of assessment by the entire sample of judges.

6.1.2 Choices based on the visceral assessment

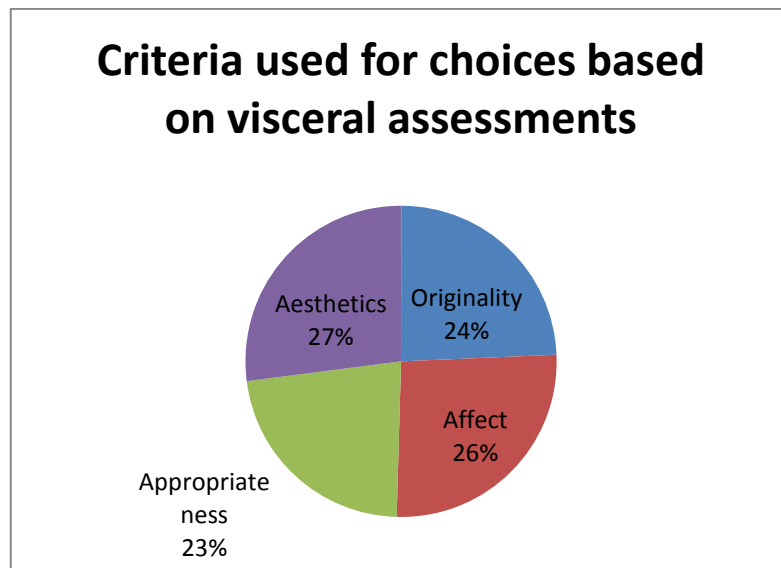


Figure 7: Percentages of occurrences of criteria on which judges base their visceral assessments.

Originality	27
Aesthetics	30
Layout	13
D-Elements	10
Quality	6
Simplicity	1
Appropriateness	29
Comprehension	16
Relevance to the subject	13
Affect	25
Emotions	9
Appeal	16

Table 24: Criteria and ERs used by judges to make choices based on visceral assessment

Figure 7 shows the distribution of criteria used by judges to explain their **choices of the most creative posters, based on the visceral assessment**.

As we can see at the Figure 7, the four criteria are used with an almost equal frequency.

A closer look to ERs (Table 24) reveals that, compared to general data collected during the whole interview, during the visceral assessment, judges talked less about the

details, such as *Design elements* and more about the general *Originality*, *Comprehension of the message* or *Appeal*.

6.1.3 Choices based on reflective assessment

Our second goal was to find out **on which criteria are based choices made after a reflective assessment**. To analyse this, we took into account the criteria for the posters chosen as the most creative after the reflective assessment stage. We took into account

only the criteria, which, during the reflective assessment, resulted with positive statements (since choices are normally based on positive observations).

The figure 8 illustrates that *Aesthetics* was the most frequently cited criterion that could explain choices of the most creative posters after the reflective assessment. *Originality* seems to be the less important criterion for this phase.

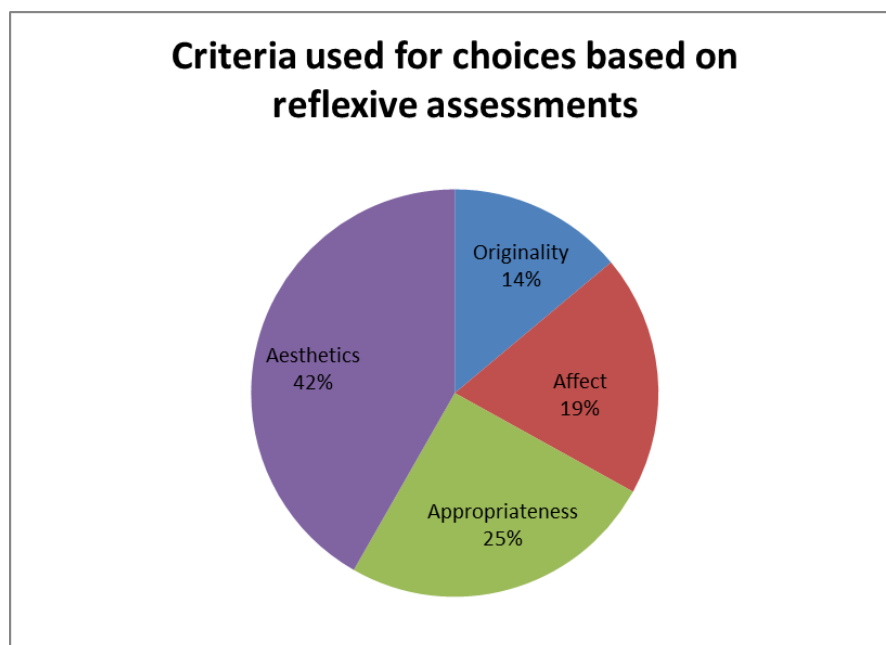


Figure 8: Percentages of occurrences of criteria resulting with positive statements on which judges based their choices after reflective assessments.

Originality	32
Concept development	16
Innovation	14
Risk-taking appreciation	2
Aesthetics	96
Aesthetics appreciation	6
Design elements	31
Quality of execution	6
Layout	35
Simplicity	13
Style	5
Appropriateness	58
Relevance to the subject	29
Compreh. of the message	29
Affect	44
Appeal	18
Emotions conveyed	26

Table 25: Criteria and ERs used by judges to make choices based on reflective assessment.

We can also observe that ERs the most frequently cited during this phase are *Layout*, *Design elements*, *Relevance to the subject*, *Comprehension of the message* and *Emotions conveyed*. It seems that positive statements concerning these criteria were the most favourable for considering a poster as creative, after a reflective assessment.

To complete this research with additional data, we selected 17 posters that were chosen on the base of visceral assessment but rejected after the reflective assessment phase. By verifying which criteria were used to assess these posters, but which resulted with negative statements, we could have a more precise idea on criteria

that appear not important during the visceral assessment, but make that are decisive during choices based on the reflective assessments.

Results showed that *Aesthetics* represents 38% of these criteria, *Appropriateness* 27%, *Originality* 22% and *Affect* 3%. Therefore, *Aesthetics* and *Appropriateness* are the criteria that in 65% of cases could explain the rejection of choices made during the visceral assessments.

6.1.4 Choices based on personal preferences

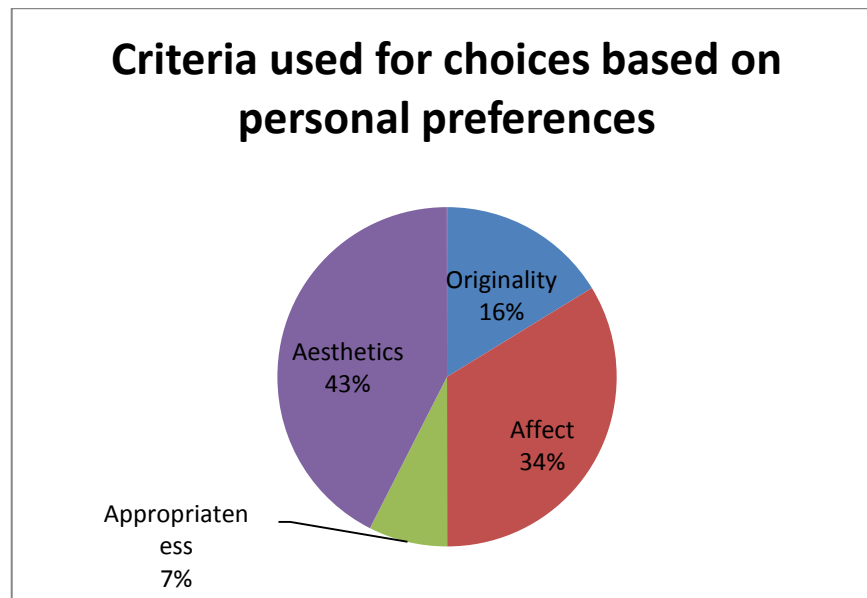


Figure 9: Percentages of occurrences of criteria on which judges based their personal choices.

Originality	13
Concept development	1
Innovation	10
Risk-taking appreciation	2
Aesthetics	34
Aesthetics appreciation	8
Design elements	9
Quality of execution	4
Layout	8
Harmony	1
Simplicity	3
Style	1
Appropriateness	6
Compreh. of the message	6
Affect	27
Appeal	10
Emotions conveyed	17

Table 26: Criteria and ERs used by judges to make choices based on personal preferences.

Finally, we wished to find out which criteria were used for the graphic design assessment based on personal preferences, when the judges are not asked to make their assessments based on the creativity or on the brief-appropriateness of the posters, but only on what they appreciated.

Figure 9 shows that the most frequently used criteria were *Aesthetics* and *Affect*. The *Appropriateness* criterion was much less frequently cited.

From the comparison of ERs' number of occurrences, we can assume that *Emotions conveyed* were the most frequently cited, followed by *Appeal*, *Originality* and *Design elements*. These four ERs cover 57% of the total criteria used to assess the posters during this stage of the interviews and it seems that they are the most related with the personal preferences of the judges.

6.2 Conclusions

First of all, the results inform us on how detailed the judges' analyses are concerning the technical details of *Design elements* and confirm the high granularity of the *Aesthetics* criterion, stated in the previous stage of this study. The fact that most judges entered deeply into analysing fonts, illustration quality and layout, given that the ERs included into the *Aesthetics* criterion occupy 43% of all ERs used during our interviews. Nevertheless, we can state that there are several ERs used more than others within each criterion category, so we could assume them to be the most representative ERs for those criteria. It seems that the major part (67%) of assessments concerning the *Aesthetics* criterion is about the *Design elements* and *Layout*. For the *Appropriateness* criterion, this role would be attributed to the *Comprehension of the message* ER, which not only takes 62% of parts in ERs within this criterion, but also is the second most-used ER amongst those listed in this study. *Originality* seems to depend mostly on *Idea innovation* (68%), while *Affect* is mostly about the *Emotions conveyed* (66%).

Originality is less discussed than *Aesthetics*, unlike in the results based on declaration of people's mental representations (see Chapter 4). This might be the consequence of the abstract nature of this criterion, which is somehow opposite to the aesthetic details that are physically present in the poster. *Aesthetics* ERs are mostly concrete elements of the posters and there are simply more opportunities to focus on them than to state such a characteristic of the whole design, as its *Originality*.

Quantitative data show us that *Appropriateness* also seems to be more discussed than *Originality*. This might be due to the brief, which induces more possibilities of references between a design outcome and multiple specific constraints, while the *Originality* leads to a single statement, which compares a design outcome with the entire domain.

When comparing the distribution of criteria and ERs in visceral and reflective assessments, we can state that on the visceral level, the criteria are distributed in a rather equal way, while on the reflective level, we observe a broad domination of *Aesthetics*.

It seems that on the **visceral** level, the creativity of graphic design is mainly based on the *Innovation* ER. The *Originality* of an outcome is balanced by its *Appropriateness*, as perceived in terms of *Comprehension of the message*. Moreover, during visceral assessment, judges seem to be sensitive to the design's *Appeal*, which can appear logical, since what we are measuring here is the first impact of the product.

During **reflective** assessment, judges appear to switch focus from the general approach (present on the visceral level) to the aesthetic details of the outcomes (mainly *Design elements* and *Layout*). Also both ERs linked to *Appropriateness* are considered to a relatively high degree. While considering *Affect*, judges switch their attention from *Appeal* to analysing the conveyed *Emotions*, within *Originality* (which becomes the less-mentioned criterion) rather than on general *Innovation*, they focus on *Concept development*.

The table below (Table 27) offers an overview of ERs that are used the most for two types of assessments: visceral and reflective assessments. We can observe a switch of focus within the four criteria from one type of ER to another. The most radical changes appear within *Originality* and *Affect*. It could be possible that judgments of *Innovation* and *Appeal* are more typical while design creativity is assessed in the very first approach, while *Concept development* and *Emotions conveyed* are considered as judges enter in greater depth into the detailed analyses and make their choices after considering their knowledge and values about creativity.

Table 27: Comparison of creativity ERs used for visceral and reflective assessments

Criterion	Creativity ERs used for visceral assessment	Creativity ERs used for reflective assessment
Originality	<u>Innovation</u>	Concept development (+ Innovation)
Affect	Appeal	Emotions conveyed
Appropriateness	Comprehension of the message (+ Relevance to the subject)	Comprehension of the message (+ Relevance to the subject)
Aesthetics	Layout + Design elements (+ Quality)	Layout + Design elements (+ Simplicity)

If we compare criteria and ERs of judges making choices on creativity and on personal preferences (both should be qualified as reflective assessments, since made after the detailed analyses of all the posters), we note that, while *Aesthetics* and *Originality* preserve similar proportions, *Appropriateness* decreases and leaves its place to the *Affect* criterion (see Table 28).

It seems that the personal preference choices are based on the *Emotions conveyed* by design. This is visible in no other type of assessment. These judgments, even after a long reflective evaluation of each poster, seem to be much less based on ERs that are in some way connected with reflective analysis: those linked to *Appropriateness* are much less frequent (which can also be connected to the fact that for personal preferences there is no importance of constraints included in the brief) and *Concept development* was replaced by *Innovation*. Overall, the most-used ERs are connected with *Emotions*, *Appeal*, *Innovation* and some elements of aesthetics (*Layout*, some *Design elements* and general *Aesthetic appreciation*).

Table 28: Comparison of creativity ERs used for assessments of creativity and those based on personal preferences

Criterion	ERs used for assessment on creativity	ERs used assessment on personal preferences
Originality	Concept development (+ Innovation)	Innovation
Affect	Emotions conveyed	<u>Emotions conveyed</u> (+ Appeal)
Appropriateness	Comprehension of the message (+ Relevance to the subject)	-
Aesthetics	Layout + Design elements (+ Simplicity)	Layout + Design elements + Aesthetic appreciation

7 Third stage of the study: criteria used by judges depending on their backgrounds and their levels of experience in design

At this stage, we wish to find out the differences in the use of criteria and ERs, depending on the judges' different backgrounds and levels of experience. Furthermore, we take into account these differences within the specific types of creativity assessments: the visceral and the reflective assessments, as well as the assessments based on the judges' personal preferences.

This part of the study, similarly to the previous parts, has an explorative character, since it is based on qualitative data. Their irregular distribution makes it difficult to truly compare the criteria used between judges with different profiles. Nevertheless, with the number of occurrences that we have access to for each criterion, and their detailed descriptions, we can produce some statistics that would give us some additional insights to complete the previous studies.

As in previous studies, we presume that in line with researchers such as Norman (1988) or Glăveanu, (2010), during a real situation of design creativity assessment, judges will use criteria differently, depending on their backgrounds and their levels of experience in design.

7.1 The distribution of judges' assessments with regard to their different profiles

In order to find out which groups of judges seem to proceed similarly within their creativity assessments, we examined the Multiple Correspondence Analysis of the criteria distribution across the different judges' profiles.

First, we compared the judges' profiles depending on their professional backgrounds; next, we performed the same analysis considering their levels of professional experience in design. The comparison included only the general results, since for more detailed analyses, we proceeded with another approach, which will be presented in the next part of this thesis.

7.1.1 Results depending on judges' backgrounds

The Multiple Correspondence Analysis allowed us to observe the preferences that some groups of judges have for using the specific criteria; Figure 10 illustrates these differences depending on the judges' backgrounds. We can observe that the group of laypeople shows a different approach in their use of assessment criteria, compared to the three other groups.

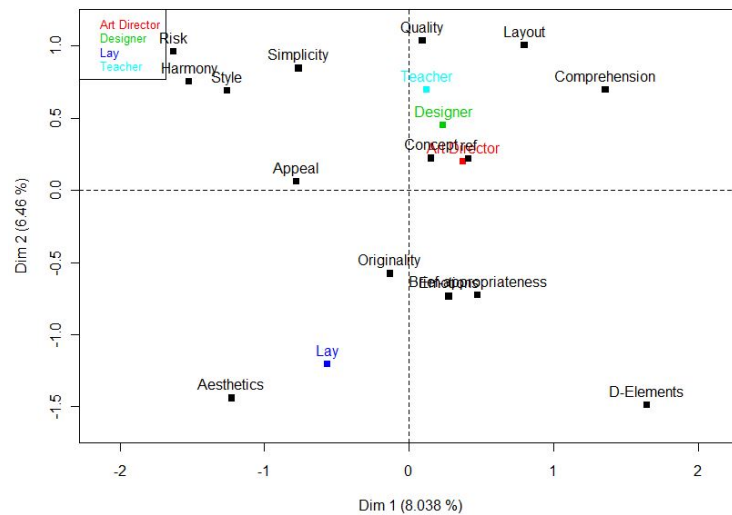


Figure 10: Multiple Correspondence Analysis of the distribution of criteria used during the whole interview, across the judges' backgrounds.

7.1.2 Results depending on judges' levels of experience

The Multiple Correspondence Analysis revealed that, again, laypeople showed a different approach than asserted experts and intermediary experts (Fig. 11). It seems that judges with some experience in the field of design have more specific preferences for assessment criteria use than those who have no experience in design at all.

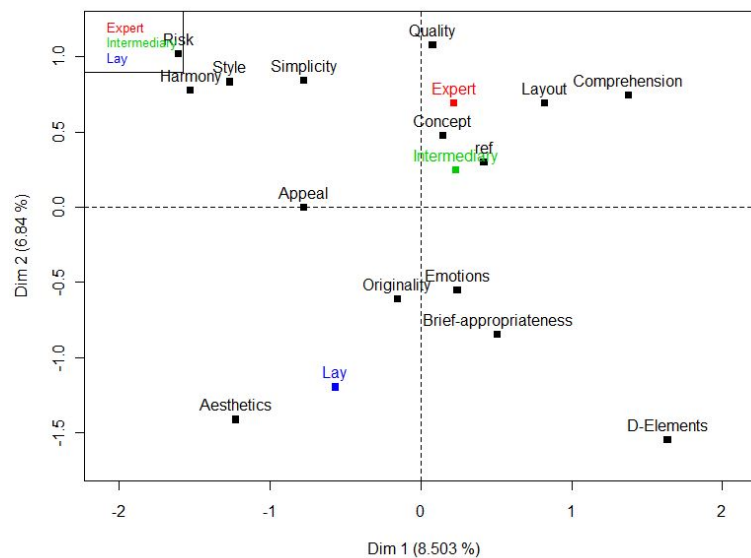


Figure 11: Multiple Correspondence Analysis of the distribution of criteria used during the whole interview, across the judges' levels of experience.

7.2 The distribution of the criteria and ER use, within the assessments of judges with different profiles

Our goal was also to know which criteria are the most used by different groups of judges, depending on their professional backgrounds and their different levels of experience in design. For this issue, first we compared the numbers of occurrences for all the criteria within all the group of judges, second, we analysed the ERs corresponding to each criterion.

In order to analyse the differences in the frequency of criteria use, appearing between groups of judges, we proceeded with Correspondence Analysis. We used the z-scores that express how far a value is from the population mean, and expresses this difference in terms of the number of standard deviations by which it differs (Kirkwood & Sterne, 2003). This analysis allows us to know which criteria / ERs are used by a group of judges with a specifically high or specifically low frequency, compared (1) to other criteria used by this group and (2) to other groups. For example, if a z-score for Aesthetic criterion for designers is of 2.4, it means that this criterion was cited on average by designers 2.4 standard deviations above the whole population's average use of this criterion. We had to choose a threshold that would mark the lower and the higher limit, above which we can consider that the frequency of occurrence of the criterion for a group is interestingly different. There is no established threshold that determines which results should be taken into account so we were free to decide if we prefer to have more or less wide range of data. We decided to the threshold of > 1.5 and < -1.5 , since we wished to find only the most important differences.

We proceeded with this analysis by comparing judges (1) with different backgrounds and (2) with different levels of experience. In both cases, we analysed their results across all the types of assessments.

7.2.1 Results depending on judges' backgrounds

The results of our analyses are presented in the tables 29 and 30. The details presenting the graphic illustration of the distribution of ERs (Table 30) and of criteria (Table 29) use.

Each of the tables presented above summarize results for four different types of assessment: first, a comparison of the global use of criteria / ERs during the whole interviews, next, a comparison of the criteria / ERs use for choices based on visceral assessment, on reflective assessment and on personal preferences. The differences between judges with different backgrounds were calculated within each type of assessment separately, but we present them grouped together, in order to have a better overview of the characteristics of their assessments.

The analysis of the **designers'** use of criteria shows us that this group of judges is especially focused on *Aesthetics*. Moreover, if we look at the ERs, we can see that this emphasis is especially placed on *Quality of execution* and *Layout*. Furthermore, this tendency is especially visible when they proceed to make creativity choices based on visceral assessment.

This group shows less consideration than others for *Innovation* and *Relevance to the subject* ERs on the visceral level, which has an impact on their use of *Originality* criterion for this type of assessment. Moreover, in general, designers quoted less *References*, compared to other groups.

Table 29: Criteria most and least used by participants with different backgrounds, for choices based on different types of assessments.

The most used criteria (z-scores > 1.5) are presented in green.

The least used criteria (z-scores < 1.5) are presented in red.

	Designers			Art Directors			Teachers			Lay		
	Visc.	Refl.	Pers.	Visc.	Refl.	Pers.	Visc.	Refl.	Pers.	Visc.	Refl.	Pers.
Originality	-1,47			0,17			0,75			0,86		
	-1,94	-0,24	0,61	1,58	0,99	1,40	-0,22	0,40	1,00	1,46	-1,28	-2,24
Aesthetics	2,04			1,92			0,04			-4,51		
	2,57	0,22	-1,07	0,32	1,17	-0,98	-0,85	-0,19	-0,15	-2,65	-1,38	1,60
Appropriateness	-1,26			-0,42			0,94			0,91		
	-1,20	0,30	-0,99	-1,07	-0,21	0,34	1,49	0,39	2,71	0,85	-0,44	-1,52
Affect	-0,28			-2,80			-1,97			5,34		
	0,50	-0,47	1,25	-0,84	-2,33	-0,03	-0,44	-0,51	-1,81	0,47	3,63	0,48

Concerning **Art directors**, it seems that similarly to designers, they focus on the *Aesthetic* criterion, but it seems to be more connected with the *Style* ER; and for choices based on reflective assessment, they seem to have a special consideration for the *Layout* ER. Their visceral choices are marked by the *Originality* criterion and by the *Originality* ER, while their reflective choices, by *Risk-taking appreciation*.

This group of judges seem to pay less attention than others to the *Impact on audience* criterion, especially during reflective assessment. An analysis of ERs shows that this is especially stated for the *Emotions conveyed*. Moreover, when it came to their personal preferences, this group showed less interest for *Design elements* than other participants.

The **design teachers'** preferences are more visible within ERs analyses: they seem to be especially marked by *References* and *Concept development*, especially when it comes to choices based on reflective assessments. When making choices based on their personal preferences, they are characterised by their consideration for *Risk-taking appreciation*. Similarly to art directors, they pay less attention to the criterion of *Impact on audience*, but in their case it is a consequence not only of a low focus on *Comprehension of the message*, but also on *Appeal*. This group seems, in general, to give less attention to the *Aesthetic appreciation* ER.

The **laypeople**, compared to others, show much more consideration to the *Impact on audience* criterion, which is visible as well from analysis of the general use of criteria, as from analysis of criteria that affect the most the choices that are based on reflective assessment. A closer look at ERs shows that group expressed the *Impact on audience* criterion by talking about both: *Appeal* and *Emotions conveyed*.

The *Aesthetics* criterion seems to take a less important place in assessments made by laypeople, with the exception of assessment situations based on their personal preferences, in which *Aesthetics* seem to play an important role. Within ERs that could inform us more about the detail of this criterion, we can see that general *Aesthetics appreciation* is more frequently used by this group than by others, as well for the global results, as for choices based on reflective assessment and on personal preferences. *Design elements* seem to be cited by this group only during choices based on personal preferences, but not during

visceral choices. Moreover, this group spoke less than others about the *Layout* and about the *Quality of execution*.

Table 30: ERs most and least used by participants with different backgrounds, for choices based on different types of assessments.

The most used ERs (z-scores > 1.5) are presented in green.

The least used ERs (z-scores < 1.5) are presented in red.

	Designers			Art Directors			Teachers			Lay		
	Global			Global			Global			Global		
	Visc.	Refl.	Pers.	Visc.	Refl.	Pers.	Visc.	Refl.	Pers.	Visc.	Refl.	Pers.
ORIGINALITY												
Innovation	-1,26			0,78			0,06			0,60		
	-1,94	0,92	1,08	1,58	-0,13	1,47	-0,22	-1,23	0,09	1,46	0,45	-1,97
Risk taking ap.	0,14			0,08			1,32			-1,69		
		-0,69	-0,57		1,78	-0,72		-0,68	2,65		-0,68	-0,88
Concept dev.	0,14			-0,04			2,32			-2,70		
		-0,92	-0,40		0,96	1,44		1,75	-0,43		-1,91	-0,62
AESTHETICS												
Aesthetics ap.	-1,49			0,55			-2,90			4,35		
		-1,19	-1,14		0,13	-1,45		-1,17	-1,22		2,24	2,78
Style	0,36			1,52			-0,40			-1,49		
		-0,17	-0,40		1,19	1,44		-0,13	-0,43		-1,07	-0,62
D-Elements	0,56			-1,20			-0,11			0,68		
	1,14	1,34	-1,21	0,47	0,50	-1,54	-0,28	-0,79	-1,30	-1,53	-1,16	2,95
Harmony	0,14			-0,51			1,32			-1,10		
			-0,40			1,44			-0,43			-0,62
Layout	2,50			0,57			-0,25			-3,15		
	2,29	-0,45	1,49	-0,64	1,93	-1,45	-0,65	0,00	1,22	-1,75	-1,77	-0,62
Quality	1,53			1,15			-0,45			-2,41		
	0,48	1,32	-0,81	1,22	-1,35	0,93	-0,38	1,39	1,44	-1,19	-1,17	-1,24
Simplicity	-0,73			1,44			0,08			-0,69		
	1,01	-0,62	-0,70	-0,38	-0,49	1,37	-0,49	0,01	0,58	-0,48	1,17	-1,08
APPROPRIATENESS												
Relevance	-0,94			-0,70			0,51			1,20		
	-1,77	0,43		-1,37	0,05		1,03	0,14		2,27	-0,63	
Comprehen.	0,14			0,02			-0,18			0,02		
	-0,02	0,05	-0,99	-0,20	-0,28	0,34	1,07	0,14	2,71	-0,90	0,14	-1,52
EFFECT ON AUDIENCE												
Appeal	-0,03			-1,21			-1,72			3,15		
	0,38	0,35	1,08	-0,20	-1,06	-1,00	0,05	-0,55	-1,37	-0,39	1,42	1,08
Emotions conv	0,23			-1,99			-1,65			3,56		
	0,32	-0,87	0,74	-1,14	-2,11	0,73	-0,80	-0,39	-1,23	1,30	3,71	-0,23
Ref	-2,26			-1,39			3,46			0,14		
		-0,36			-0,61			2,24			-1,17	

We observed that laypeople take *Originality* and *Appropriateness* less into account, during assessments based on their personal preferences. Within the *Originality* ERs, we can assume that they seem to talk less about *Risk-taking appreciation* and *Concept development*, compared to other groups. When considering *Appropriateness* ERs, it seems

that *Comprehension of the subject* is not important for the personal preferences of this group. Nevertheless, it seems that *Relevance to the subject* is considered during the visceral assessments of this group.

7.2.2 Results depending on judges' levels of expertise

We carried out the same analysis, considering differences of the judges' levels of experience. The details presented in tables 31 and 32 illustrate the distribution of ERs (Table 32) and of criteria (Table 31) use. Again, we used the z-scores with the threshold results > 1.5 and < -1.5 for four different types of assessment.

Table 31: Criteria most and least used by participants with different backgrounds, for choices based on different types of assessments.

The most used criteria (z-scores > 1.5) are presented in green.

The least used criteria (z-scores < 1.5) are presented in red.

	A. experts			Interm. experts			Lay		
	Global			Global			Global		
	Visc.	Refl.	Pers.	Visc.	Refl.	Pers.	Visc.	Refl.	Pers.
Originality	0,04			-0,62			0,86		
	-0,74	1,59	1,53	-0,43	-0,41	-0,11	1,43	-1,22	-1,07
Aesthetics	0,10			2,95			-4,51		
	-0,92	-0,33	-0,86	2,61	1,33	-0,62	-2,62	-1,44	1,20
Appropriateness	0,35			-0,92			0,91		
	0,21	0,13	2,86	-0,83	0,26	-0,68	0,93	-0,50	-1,60
Affect	-0,61			-3,05			5,34		
	1,50	-1,00	-1,44	-1,46	-1,97	1,09	0,34	3,81	0,15

We can see that asserted experts are focused on *Originality* criteria when it comes to choices based on reflective assessments and personal preferences. They seem especially attentive to Risk-taking, *Concept development* and *Style* and using *References* seems to be important for their reflective choices. For choices based on their visceral assessment, they seem to be especially influenced by the *Impact on the audience*, while their choices based on personal preferences seem to be more dependent on *Appropriateness*, and more specifically on *Comprehension of the subject*, but less influenced by *Appeal* or *Design elements*.

Intermediary experts seem to focus on *Aesthetics*, globally and when making choices based on visceral assessment. Within the *Aesthetics* criterion, they seem to consider the ERs, like *Design elements* and *Layout* (especially for visceral assessments) and *Style* (for reflective assessments), but not the general *Aesthetics* appreciation. On the other side, they seem to take less into account the *Impact on audience* criterion, which comprises both ERs: *Appeal* and *Emotions conveyed*.

In comparison to professionals, Lay people's results are very similar to those, where the comparison was made on participants' backgrounds.

Table 32: ERs most and least used by designers with different levels of expertise in design for choices based on different types of assessments.

The most used criteria (z-scores > 1.5) are presented in green.

The least used criteria (z-scores < 1.5) are presented in red.

	Asserted Expert			Intermed. Experts			Lay		
	Global			Global			Global		
	Visc.	Refl.	Pers.	Visc.	Refl.	Pers.	Visc.	Refl.	Pers.
ORIGINALITY									
Innovation	-0,84			0,08			1,04		
	-0,74	-0,23	0,32	-0,43	-0,08	0,42	1,43	0,38	-0,61
Risk taking ap.	1,14			0,13			-1,71		
		0,45	2,12		0,12	-0,81		-0,69	-0,92
Concept dev.	0,89			1,20			-2,74		
		2,13	1,50		-0,48	-0,57		-1,82	-0,65
AESTHETICS									
Aesthetics ap.	-1,44			-1,83			4,27		
		-1,39	-1,41		-0,40	-1,61		2,16	2,49
Style	1,68			-0,60			-1,51		
		-1,13	1,50		1,67	-0,57		-0,97	-0,65
D-Elements	-0,57			1,62			-1,29		
	-0,58	-0,29	-1,50	1,56	1,15	-1,71	-1,54	-1,23	2,65
Harmony	-0,16			1,05			-1,12		
			-0,50			1,18			-0,65
Layout	1,00			0,42			-1,89		
	-0,96	0,24	0,71	2,03	1,14	0,25	-1,75	-1,84	-0,76
Quality	1,03			0,79			-2,41		
	0,41	0,32	0,00	0,45	0,53	1,49	-1,09	-1,09	-1,30
Simplicity	1,33			-0,85			-0,73		
	-0,55	-0,08	0,29	0,79	-0,73	1,04	-0,49	1,09	-1,13
APPROPRIATENESS									
Relevance	-1,19			1,55			-0,35		
	-0,84	0,01		-1,09	0,45		2,47	-0,63	
Comprehen.	0,01			0,06			-0,08		
	1,00	-0,09	2,86	-0,15	0,04	-0,68	-0,92	0,05	-1,60
EFFECT ON AUDIENCE									
Appeal	0,14			-2,50			2,98		
	1,28	-0,32	-1,58	-0,67	-0,70	0,42	-0,51	1,33	0,85
Emotions conv.	-1,14			-1,49			3,44		
	0,79	-1,24	-0,61	-1,55	-1,94	1,05	1,28	4,09	-0,46
Ref	0,50			-0,87			0,43		
		2,22			-1,02			-1,19	

The analysis of the **asserted experts'** use of the criteria shows us that in general this group of judges is especially focused on *Style*. They do not show any specificity during the visceral-based choices, but we noted that within their reflective-based choices their consideration for *Concept development* and for *References* is higher than for other groups. During the personal preferences-based choices they consider more *Comprehension of the message*, *Concept development*, *Risk-taking appreciation* and *Style*, but they give less attention to *Appeal* and *Design elements*.

General assessments of **Intermediary experts** seem to be especially marked by *Design elements* and *Relevance to the subject*. Their visceral choices seem to be based on *Layout* and *Design elements*, while the reflective ones seem to be based more on *Style*.

Nevertheless this group seems to give less attention to *Appeal* and *Aesthetics* (the last one also especially less taken into account within personal preferences). They use *Emotions conveyed* less than other groups did, during both visceral and reflective choices. Moreover, even if *Design elements* seems to be important to them during the creativity assessments, when comes to the personal preferences, they do not base their choices on this criterion.

Compared to the experts' groups, **Laypeople** show more consideration to *Emotions conveyed* (also when making reflective-based assessments) and *Aesthetics* (also for reflective- and personal preferences-based assessments) and *Appeal*. During their visceral choices, they seem to be especially focused on *Relevance to the subject*, while their choices made on the personal preferences are based especially on *Design elements* and *Aesthetics*.

In general, they seem to take less into consideration the *Layout*, *Concept development*, *Quality of execution*, *Style* or *Risk-taking appreciation*. *Design elements*, even if they are taken into account during the personal preferences based choices, do not seem to have such an impact on visceral-based choices. When making choices on their personal preferences, they attach less importance to *Comprehension of the message* than other groups do.

7.3 Conclusion

In the last study we did not focus on the frequency of occurrence of each criterion within each group of judges, but on the tendency for some judges' profiles to quote a specific criterion more than average. Our goal was to verify if participants with different profiles showed an inclination to focus on different criteria.

We could observe that, these tendencies are not the same, depending on the judges' background and on their level of experience.

More specifically, two differences seem to be the most salient, when we look at the results. Both involve laypeople and professional designers, especially those who have an intermediary level of expertise in design. These differences are also visible, when comparing the results of participants with different backgrounds. The first one was observed between designers and art directors on the one hand and laypeople on the other hand. The two first groups seemed to be highly focused on the **Aesthetics** criterion while the laypeople seemed to be the group less interested by this criterion. The other opposition seems to concern the **Impact on audience** criterion, which seems to be in the center of interest for laypeople, but not for gatekeepers (art directors and teachers), who cited it less than others.

We could believe the participants with at least some expertise in design to be more focused on *Originality*. It would be especially expected from designers, in line with the assumption made by Norman (2004) that in their professional field, they wish their work to be appreciated not only as "pretty", but also as resulting from deeper processes of creative work. In this study, designers seem to talk especially about the *Quality of execution* and the *Layout* of the assessed artifacts. Therefore, we should interpret these results as an analytical approach to creativity. Unlike laypeople, who are focused on rather general comments about Aesthetics that could be classified as the concentration on what is "pretty", designers seem to be interested in the aesthetic details. This could result from the

fact that professionally they deal with these detailed problems to obtain the outcomes that would be appreciated as creative.

We observed some more attention given to *Originality* coming from the profiles representing the gatekeepers, detectable within the specific ERs. Nevertheless, these *Originality* ERs are not the same for the two profiles of gatekeepers. Art directors seemed to be more focused on *Innovation* when making choices based on the visceral assessment of creativity and they appreciated the *Risk-taking* while making choices based on reflective assessments. We could interpret this in line with their gatekeeper profiles, that comprises a large domain knowledge and experience in evaluating designers' work (Hooke, Nakamura & Csikszentmihalyi, 2003). In a first approach (during visceral assessment), an art director is quickly distinguishing the traits of *Innovation* that are present in the poster, which, upon closer inspection (reflective assessments) is completed by the consideration of the risk taken by the author to express this innovation, as well as of his or her *Style* (another ER considered by this group to a higher than average degree). We could conclude that for art directors, *Originality* assessment relies on their detailed knowledge about existing trends in graphic design and borders that should be reached (but not exceeded) by creative individuals, to obtain the outcomes judged as original.

The other group of gatekeepers, design teachers, seemed to be especially focused on *Concept development*, which could be interpreted as a consequence of their professional consideration of the creative process rather than the product itself (Birenbaum, 2007). They use also more *References* than others, which is probably the consequence of their extended domain knowledge (Amabile, 1996). Their will to use multiple examples and analogies (within the design domain, but also external to this domain) to illustrate and support their discourse, could reflect the way in which they share their knowledge with students. To express their personal preferences, teachers seem to appreciate the *Risk-taking* and the *Comprehension of the subject*, which are directly connected with the two fundamental criteria commonly cited in different definitions of creativity: *Originality* and *Appropriateness* (Sarkar & Chakrabarti, 2008). It seems that these criteria are strongly internalized by this group of participants, since they use them to express purely personal preferences, concerning not even creativity, but individual taste.

Furthermore, art directors and teachers seldom mentioned the emotional aspects of the designs. This is surprising, especially when observed in the assessments of art directors, who are supposed to know how to attract the audience's attention and increase the effectiveness of the graphic design (Csikszentmihalyi, 1999). On the other hand, laypeople clearly show an approach opposite to gatekeepers, since they widely cited the *Impact on audience* criterion and the ERs like *Appeal* and *Emotions conveyed*, which would be in line with the model of Horn and Salvendy (2006b), in which users perception of design creativity is first of all influenced by their affect. Moreover, they speak less about the aesthetic details like *Layout* or the *Quality of execution* than participants with higher level of experience in design. It could be a consequence of their less analytic approach and focus on general feelings rather than on the technical details. Rather than criticizing, for example, the way in which the title's font was placed in relation to the main illustration (typical comments made by the professionals), these participants were simply saying that they found the poster ugly and sad, without being able to explain why. This could be the consequence of the low domain knowledge and the lack of professional vocabulary that does not allow them to focus on details appreciated by design experts. There is a single case of assessment in which laypeople seem to switch their focus to the design aesthetic details. While the professionals (especially the intermediary experts) talk about the details

during the creativity assessments, the laypeople start to use them rather when talking about the personal preferences.

Nevertheless, we cannot say that laypeople react only to emotional signals. Their visceral assessments are based on emotional ERs to the same degree as those of design professionals. Moreover, within visceral assessments laypeople commented the aspects related to the *Relevance to the subject*, surprisingly more frequently than in the case of other groups, which could lead us to conclude that they seem to focus on the functional side of the designs, which would be in line with findings described by Wojtczuk and Bonnardel (2012) in the field of product design. According to them, non-designers' (or users') mental representations of objects are based more on their functionality, since their only previous experience with design is based on its use, not on creative activity. Similarly, in the field of graphic design, laypeople (or the potential audience) are mainly focused on the functional part of posters, like readability of the most important information, or the immediate understanding of the message.

Another interesting point concerns the difference observed between the asserted experts' and intermediary experts' results. Aesthetics seems to be more important to intermediary designers, especially for choices based on visceral assessments. The most important ERs seem to be *Design elements*, Layout and Style. Asserted experts do not seem to pay excessive attention to this criterion (except for the Style ER). The asserted experts' focus seems to be placed more on the Impact on audience, during the visceral assessment, which compared with the laypeople (or potential audience's) preferences, gives the better image of them as those who understand the audience's needs and interests.

Asserted experts also seem to focus more on *Originality* (mainly on Concept development), during the reflective assessment. It seems that less experienced design professionals focus more on technical aspects of design outcomes, which could be interpreted as focusing on the means used to achieve the final effect. Instead, the more experienced practitioners share the audience's preference for "big picture" approach, even if in case of asserted experts this general view is influenced by the domain knowledge, while the audience's approach is influenced by their implicit theories about creativity, self-related interpretations and everyday knowledge (Leder, Belke, Oeberst & Augustin, 2004). It seems that asserted experts evaluate in the first the visceral approach the impact that the poster could possibly have on its public, while its *Innovation*, *Concept* analyze and searching for possible *References* is made only during the reflective approach.

PART III – DISCUSSION

Chapter 7 General discussion and research perspectives

The results of the three studies presented in this thesis have enabled us to identify in greater detail the creativity criteria used in the area of graphic design. Furthermore, it allowed us to understand the influence of judges' characteristics, such as their professional background or level of experience in design, on the use of those criteria.

In chapter 4, we analyzed the mental representations of creativity within this area and showed several variances that could differentiate graphic design from product design and to better understand the judges' mental representations of creativity in graphic design. Chapter 6 gives us more insight about the nature of the different evaluative referents used to assess graphic design thanks to the presence of the context in which they were expressed and of verbatim that provide some refinements to our understanding of creativity criteria.

Moreover, the results of this research allowed us to analyze differences within creativity assessments of graphic design, in order to identify some characteristics of judges with different professional backgrounds and level of experience in design. For each group of judges, we identified which criteria seemed to be the most available for heuristics-based assessments of graphic design creativity: chapter 4 describes them on the mental representations level, chapter 5 presents which criteria seemed to guide mostly the attribution of creativity scores, and chapter 6 gives some highlights on the frequency of criteria use within spontaneously expressed assessments. Moreover, we identified for which criteria the inter-judge agreement, within each group of judges (described in chapter 5), was not the same.

In this chapter, we will discuss our main results and analyze the use of ERs and criteria by judges in different contexts of creativity assessment.

1 Summary of our findings

In this thesis we used three different methods to approach the assessment of creativity within the area of graphic design.

We used a survey with open questions in order to build knowledge about the mental representations of creativity within a population of participants having different backgrounds and levels of experience in design. We asked them to tell in their own words what is the most important for them to assess creativity in the areas of graphic design and product design. In consequence, we obtained ERs that allowed us to understand the specificities of graphic design, in comparison to the other, most studied field of creative design. Furthermore, we could compare the nature and quantity of answers obtained by participants with different profiles and the results of the entire sample with those found in other research.

Next step was to switch our focus from mental models of creativity to the actual use of criteria within real-world situations of graphic design assessment. The most frequently mentioned categories resulting from the survey were tested as criteria on which judges (still with the regard to their different profiles) were asked to assign scores while assessing the creativity of graphic design outcomes. The scores assigned for specific criteria were compared with those assigned for creativity, in order to find out which of criteria were most correlated with creativity scores within different groups of judges. Moreover, inter-judge agreement was studied within all groups, for each criterion.

Finally, we mixed two previously used approaches by putting the judges in real situations of graphic design assessment, but without providing them any ready-to-use lists of criteria. The verbatim utterances were collected during three different types of assessment: Two of them focused on creativity at the visceral and reflective levels (Norman, 2004) and one on personal preferences. We analyzed the verbal utterances and adjusted the categories made during the first study (based on participants' mental representations of creativity). This study allowed us to establish some tendencies about the frequency of occurrence of the ERs within different groups of judges, during different types of assessments, but the main interest of this study is to provide qualitative data about the context in which these ERs occur.

1.1 Summary of our findings concerning the criteria and ERs used to assess the creativity of graphic design

In the first place, we confirmed our first general hypothesis presented in chapter 3 of this thesis. Chapter 4 gave us an insight into the differences between the mental representations of creativity in two different areas of design. The frequencies of occurrences for each criterion within each area of design show that, whilst for product design the *Audience-appropriateness* criterion was the most popular criterion that seemed important for creativity, it seemed to be only moderately important in the area of graphic design, since more than 50% of participants did not even mention it. It seems that the majority of our participants believe that graphic design creativity is founded on the *Originality* criterion.

These mental representations appear to play an important role in the perception of creativity in graphic design at the visceral level. When judges are asked to explain their choices of the outcomes that seem to them more creative than others, the most frequent justification (the most frequently used ER) is linked to the *Originality* of the chosen design. Nevertheless, the later studies showed that while many participants declared *Originality* to be important, when it comes to analyzing real design outcomes via free

verbalizations, this criterion is not as much used as, for example *Aesthetics* or *Appropriateness*.

The analyses of the verbalizations presented in chapter 6 allow us to understand more in detail how creativity judges approach *Originality*. The results of this study revealed that this criterion can be related not only to the outcome, but also to its designer's ability for both risk taking and conducting a full conceptual process to defend his or her idea. It seems that the degree of innovation of the outcome is considered the most and especially in situations when the decision has to be taken quickly. Nevertheless, when it comes to deeper, reflective analyses, the *Originality* might be perceived not only at the level of *creative product*, but also at those of *creative person* and *process* (Rhodes, 1961; Hasirci & Dermirkan, 2003).

Aesthetics seem to be the mostly mentioned criterion, during our study of verbalizations, nevertheless, according to the qualitative analyses this frequent occurrence does not imply importance for creativity assessments. The high frequency of its occurrence is probably due to the fact that this criterion has the highest granularity (it is composed of numerous sub-criteria). Judges have lots of things to say about design elements (e.g. colour, layout, fonts, etc.), we could even suppose that they evaluate *Originality* and appropriateness of each aesthetic element separately. Nevertheless, the experimental study presented in chapter 5 revealed that seldom does the Aesthetics criterion guide the way in which judges attribute creativity scores (only Art directors' creativity judgements seemed to be influenced by this criterion). It seems that somehow judges are able to separate conceptual creativity from aesthetic creativity. It might be influenced by the fact that in our study, judges knew that they were assessing students' work and felt bound to be more indulgent on technical imperfections.

The ***Brief-appropriateness*** criterion was identified in chapter 4, similarly to *Originality*, as the one that differentiated graphical design from product design. It is highly considered as important for graphic design creativity, but for the product design creativity this criterion was not even mentioned by 30% of participants. The criterion seems important for creativity in people's mental representations and in their way of attributing scores on creativity. Nevertheless, it was observed during the measures of inter-judge agreement presented in chapter 5, as well as during the qualitative analyses of verbatim utterances presented in the first part of chapter 6, that judges highly disagree when assessing graphic design according to this criterion. We proceeded to change the categorisation of our two appropriateness criteria (*Brief-* and *Audience-appropriateness*), which appeared to be clearly different from each other while studied as participants' mental representations, based on declared criteria, but merged together after an in-depth analysis of verbatim utterances collected in-context. The study of mental representations, conducted outside of the context of real situation of assessment, showed that people distinguished: (1) the need for design to express what is described in the brief, from (2) the requirement to express it in a way in which it is possible for the audience to understand it. Nevertheless, in a real situation of assessment, those two aspects of appropriateness seemed to be processed at the same time. The participants were identifying their own opinion on the suitable nature of metaphors proposed by designs, with those of the potential audience, taking into account their own understanding of it as a measure of appropriateness. We may be facing here the problem observed by Lera (1981) in designers: they tend to apply their personal values to judge appropriateness, if the information in the brief is not precise enough. The conclusion that we could make here would be that if judges do not receive very precise constraints on the design subject, they would use their own constraints, which could be a major source of

variation. Graphic design uses metaphors to illustrate a subject and in our study these metaphors seemed to result in different connotations.

Even if the *Audience-appropriateness* merged with *Brief-appropriateness* during the analysis of verbalizations, we could observe within the judges' criteria that many of them were considering **Affect**. We could assume that in the specific case of graphic design the *Audience-appropriateness* is stronger connected with its emotional than with its functional aspects. These findings are in favour with those of Tractinsky et al. (2000), claiming that affect can influence the perception of function. They validate the importance of studies that take the emotional aspects of design into account (Norman, 2004; Horn & Salvendy, 2009; Demirkan & Afacan, 2012).

1.2 Summary of our findings concerning the influence of judges' profiles on creativity assessments of graphic design

1.2.1 Professional background

In line with the second general hypothesis of this thesis, we expected the judges' backgrounds and their level of professional experience in design, would have an influence on their way of assessing creativity. This hypothesis was also validated, since we could observe that across different studies, judges with different profiles were using creativity criteria to various degrees. Even if in each study the samples were not very big (which is a common problem if the targeted population represents a very specific profile of experienced professionals), we examined the same type of groups across different types of creativity assessment studies. By examining which criteria seem to be important for a specific profile across different assessment situations, we could consider that these criteria are especially important for the specific profiles.

We observed that during their creativity assessments, **designers** and **design teachers** seem to take into consideration *Originality* in particular – this is visible in their mental representations and in the way in which they attribute scores to graphic design outcomes. In the case of designers, a high inter-judge agreement is also observed for this criterion. **Art directors** are less focused on this criterion when attributing scores; yet this group mentioned it more than others, while explaining their visceral choices, which shows that art directors are also deeply affected by this criterion, but their professional background conducted them to consider rather the *Brief-appropriateness* criterion, since they seem to base on it their mental representations and creativity scores.

The importance of *Originality* criterion connects designers' and teachers' way of perceiving creativity, which is in line with the study of Wojtczuk and Bonnardel (2012). These authors used Norman's model to ground their hypothesis about the distinction between two types of backgrounds: the "creative" ones (designers and teachers) and "user-centered" ones (art directors and audience). In chapter 5, this distinction is more visible, since its results illustrated the designers' and teachers' focus on the creative quest by balancing between *Originality* and *Audience-appropriateness* criteria, while the audience and art directors seemed to focus on the informative function of the poster described in the brief. Nevertheless, while analysing all the results of our experiments, we can see that within "creative" profiles, the interest for *Audience-appropriateness* is not the same, since even if teachers show a high level of inter-judge agreement on this criterion, it appears less frequently within their mental representations of creativity and their free verbalizations. Concerning art directors and lay-people, they do not show more common points than those previously mentioned in chapter 5.

Designers and **art directors** seem to share a special focus on the **Aesthetics** criterion. This distinguishes them from teachers, who seem to be interested less in technical details and more in conceptual research. Both groups declared this criterion as being important for creativity and cited it more than other groups during their spontaneous verbalizations. Nevertheless, even if both groups consider Aesthetics as important for creativity, they mention it during their spontaneous verbalization using different evaluative elements (designers focus more on Layout and Quality of execution, while art directors focus more on Style) and, their attribution of creativity scores does not seem to be based on this criterion. We can assume that the background of those two profiles of judges makes them very sensitive to the aesthetic aspects of design, but they seem to consider them separately from the main creativity judgment.

We chose art directors and design teachers as two different profiles that represented **gatekeepers** of the design domain: teachers leading the transmission of domain knowledge, art directors taking decisive role in releasing new design outcomes into the field. Our studies show that these two types of backgrounds shape the judges' perception of creativity in different ways (described in the paragraphs above). The only common point that we found is their tendency to use certain *Originality*-based evaluative elements more than other groups do (even though teachers seem to be especially focused on the conceptual side of originality), which could reflect their large domain knowledge and experience in evaluating the designers' work (Hooker, Nakamura & Csikszentmihalyi, 2003), allowing them to easily compare a specific design with what was already accomplished in the field. Moreover both profiles have a surprising tendency to deny the emotions conveyed by design outcomes, which places them on the opposition to laypeople, whose judgments, following to Horn and Salvendy (2006b), are at the first place influenced by their affect. Therefore, we can assume that gatekeepers' judgments are made with more emotional distance, which allows them to make their creativity assessments without being influenced by elements driven by factors other than their domain knowledge.

When comparing judges with different level of professional experience in design, we observed that for both, **asserted experts** and **intermediary experts**, the *Originality* criterion appeared to be important in both contexts: as a component of their mental representations of creativity, and as being highly correlated with creativity while attributing scores to design outcomes. Moreover, both profiles showed a high inter-judge agreement on this criterion. Asserted experts seem to be even more focused on this criterion, since they mentioned it frequently to explain their choices made on reflective assessments and on personal preferences.

Intermediary experts tended to use the **Aesthetics** criterion during their verbalizations. Moreover, they seemed to consequently consider the **Audience-appropriateness** as important for creativity, since it takes a significant place in both, their mental representations and attributing scores for creativity.

It seems that the difference between the asserted experts and intermediary experts is in the quantity of the criteria that affect their representation and assessments of creativity. For both profiles the number of these criteria is higher than for laypeople, nevertheless, it seems that intermediary experts have tendency to use more criteria than participants, who accumulated more experience. This could be interpreted in line with Bonnardel (2003), who observed that less-experienced designers have more problems with managing constraints. Since the lack of sufficient experience in design is connected with constraints

management problems within the creative process, it could also influence the judges' problems with managing criteria during creativity assessments, since both processes involve the use of evaluative referents, based on the constraints defined from the brief and internal reflection and preferences (Bonnardel, 1996; Bonnardel, 2000; Bonnardel, 2006; Visser, 2009a).

In the case of **lay-people**, compared with other groups, we observed more differences between declared beliefs (mental representations) and their genuine creativity assessments. Within their mental representations based on declared criteria, this group agrees the most on *Originality* and *Aesthetics*, but during real creativity assessments we can barely tell which criterion really influences the way in which they attribute the creativity scores (the only tendency relates to *Brief-appropriateness*). It validates Norman's assumption (1988) about differences in designers' and users' mental representations of design. The lack of consistency in lay-people's use of criteria between different studies could be an effect of a great heterogeneity of this group. According to Chan and Chan (1999) and to their descriptions of implicit theories, the implicit theories built by lay-people are based on their self-related interpretations and personal experiences. Thus, if our population does not share a specific education or a particular type of work environment, their personal experiences are much more heterogeneous than those of, for example, art directors. Moreover, the mismatch of lay-people's mental representations of creativity with their attribution of creativity scores and free verbalizations, might be a consequence of the fact that, contrary to design professionals, they have neither domain knowledge, nor experience in evaluating and questioning the creativity of graphic design. Their mental representations of creativity are in line with the general definition of creativity (probably rather by an intuitive understanding than by explicit knowledge), since they mostly cite *Originality* and *Brief-appropriateness*. These mental representations might be difficult to apply while it comes to assessing real-world designs, for participants who are not used to expressing their opinion in matter of creativity.

Nevertheless, we can observe a high degree of inter-judge agreement on the *Audience-appropriateness* criterion and frequently used evaluative referents connected with *Effect on audience* within their spontaneous verbalizations. Moreover, we can observe that their reflective judgments are focused on *Affect*. Thus, we can assume that lay-people base their judgment on global impressions and emotions that they can probably hardly translate into scores.

2 Research perspectives

The interest of our study can be perceived on different levels: for the area of graphic design, precisely studied in this thesis, but also for design domain in general, as well as for other creative activities. We propose a methodology of approaching creativity in a specific field, from different perspectives, since it take into account as well the mental representations of creativity as their application in different real assessment situations. Moreover, we take into account that assessing the creativity of an outcome can be different from simply assessing an outcome using personal preferences, which are not always linked to creativity. This methodology allows obtaining data on how creativity is considered and assessed in any specific field. The main interest of results obtained with our methodology is in their ecological value, since they are based on evaluative referents that were spontaneously expressed by judges, without any suggested definitions or measures. On the other hand, these evaluative referents are confronted with the criteria already studied by

other authors and used in typically experimental methods based on scores attribution and inter-judge agreement.

For example, our results show that in product design, creativity is perceived especially on the functional level, in graphic design, originality aspects are especially taken into account. The further studies should continue this comparison by observing mental representations in both mentioned areas of design and comparing, to which degree these mental representations are applied during real situations of creativity assessment. This comparative study could be also conducted in other areas of design, like architecture, textile or digital design, in order to find out to which degree design can be considered as a general domain of research and, on the other side, how to adapt the research approach depending on its specific areas.

Moreover, in this thesis, for our studies of real assessment, we used the outcomes produced by design students. The advantage was to obtain various responses to the same design problem, which could result in broad range of positive and negative assessments on different criteria. Nevertheless, judges could adjust the criteria used to assess the unexperienced designers' outcomes, and their level of requirements toward design, to the students' level of skills. Therefore, it would be interesting to compare our results with assessments of designs created by professional designers.

Our research gives some insights about the specificities of criteria used for assessing creativity in graphic design field. We could reasonably expect *Originality* criterion to be the main source of variation in inter-judge agreement, since stating to which degree a production is different from other productions in the field, requires a wide knowledge about this field. Nevertheless, scores given in the second study, and analyses of verbalizations expressed in the third one, showed that judges mainly disagree on the *Appropriateness* criteria. Moreover, in the course of our studies, we found out that the difference between *Brief-appropriateness* and *Public-appropriateness*, even if it was clear in the participants' mental representations of creativity, becomes blurred when it comes to their natural, spontaneous verbalizations. It might be specific to graphic design area that it is difficult to distinguish the degree to which a concept presented in the brief is represented, from the degree to which it is represented in a way understandable for the audience. The use of *Appropriateness* criteria should be explored in further studies. In order to verify if judges are able to separate their own values from those of the designs' actual audience, the characteristics of the population targeted by design should be precisely specified and remain as different from the judges who participate to the experimentation as possible.

This thesis gives also some insights about the influence of judges' profiles on their creativity assessments. This multiple feedback approach can be explored in any other fields employing creative activities. The knowledge about the judges' characteristics might be helpful to choose the most appropriate group of judges (depending on the studied criteria) for studies employing consensual assessment technique, or for any other research employing the participation of judges. Following to the goals of researchers, using the specific profiles of judges gives more probability of obtaining inter-judge agreement. Moreover, our results give more qualitative information about the understanding that different groups of judges might have of creativity, while assessing it. An additional source of evaluative referents could be explored, using verbalizations exchanged by judges with different profiles during the collaborative situations of creative production and evaluation. These verbalizations could be collected during oral discussions, but could also be extracted from exchanges on electronic supports, like emails, chats or even in virtual realities

including the use of avatars with different degrees of anonymity, actually studied by CREATIVENESS program¹.

In order to collect reliable data, we tried to give as much diversity to our samples, as possible: different participants were invited to contribute to different assessment situations that we created. Nevertheless, our results could be confronted with the ones obtained with different, possibly larger samples.

Further research exploring characteristics and viewpoints of different design actors should be continued, in order to expand the knowledge about their profile characteristics. In the same line, differences between different levels of expertise in design should be explored. First, it should be determined whether the perception of creativity evolves in linear way. Second, a special focus should be put on the evolution of different criteria used. In consequence, we could know more about the degree of difficulty to use different criteria.

3 Perspectives for applied design

Understanding the assessment with the participation of judges is important as well for the assessed person, during his or her creative process, as for the institutions connected with graphic design.

In the first case, it could be helpful, for individual designers, to anticipate the judgment of their work, by taking into account the different expectations that different groups of potential design audiences might have. Until now, most studies seeking to enhance the user-centred design process have only considered the designer's and user's approach. We have extended this approach, by introducing (1) the art directors' perspective, for their role of expert decision makers on how to bring a product on the market to positively affect the audience, (2) the vision of design teachers, who shape the young designers' perception of what is important in design creativity, from the theoretical viewpoint. Anticipating different viewpoints could help designers adjust their own assessment criteria and to construct their argumentation to defend their vision, in order to communicate with different design actors in a more efficient way.

There is also a possibility of implementing the knowledge about judges profiles into the IT-based "critiquing systems" that help designers assess their own design solutions. These systems could be improved in order to reflect the perspectives of the different groups of judges. Comments and questions simulating a dialogue with a person representing one of the perspectives described in our thesis could be included. This idea could be also used as creativity-enhancing *personas*, a method currently used in ergonomics and user experience studies that aims representing hypothetical characters boasting the attributes of real users. Personas are used in user-centred design to avoid the self-centred approach (Brangier & Bornet, 2011; Pruitt & Adlin, 2006). In the case personas could not only represent the users' profiles and needs (as in the case of personas), but could also play the role of a critical partner who would question the designer about the creativity criteria of his or her project and influence their creativity during the externalization process (Wojtczuk & Bonnardel, 2010)

In the case of companies involved in graphic design creation, it can be very useful to better understand the hierarchy of importance between criteria and to anticipate potential sources of disagreement in the assessment of graphic design projects. This could be especially

¹ Research project ANR 2013 (Programme Blanc) *Activités créatives en environnements virtuels* (CREATIVENESS; Coordinateur: Lubart, T. [Université Paris Descartes]).

interesting in the context of creative collaboration. Moreover, they should be aware of differences that could appear between creativity approaches of design professionals and those of the targeted audience. If a company aims to emphasize the creative character of their products, the audience's understanding of creativity in their field should be studied, in order to give to designers a clear and detailed brief and avoid disagreements within their team.

From the educational perspective, studying the evaluative referents employed to assess design in the professional world, could be interesting to complete teachers' assessment criteria. We observed that sometimes the teachers' approach to the design differs from the approach of the design professionals. By exposing student to multiple viewpoints typically used in graphic design, it is possible to prepare them better to the professional life, but also to increase their ability to take into account different expectations.

The results of our work need to be confirmed with more samples. Nevertheless, a multiple feedback methodology, by collecting the mental representations of creativity in a specific field and confirming its validity during assessments of real outcomes, could be a way of achieving knowledge about the existing demand for creativity in any area of design.

BIBLIOGRAPHY

A

- Akin, O. (1986). *Psychology of Architectural Design*. London: Pion.
- Amabile, T. M., Conti, R., Coon, H., Lazenby, J., & Herron M. (1996). Assessing the Work Environment for Creativity. *Academy of Management Journal* 39, 1154–1184.
- Amabile, T. M., Philips E.D., & Collins M.A. (1994). Creativity by contract: Social influences on the creativity of professional artists. Unpublished manuscript, Brandeis University.
- Amabile, T. M. (1982). Social psychology of creativity: A consensual assessment technique. *Journal of Personality and Social Psychology*, 43, 997–1013.
- Amabile, T. M. (1996). *Creativity in context*. Boulder, Colo.: Westview Press.
- Andiliou A., & Murphy K. (2010). Examining variations among researchers' and teachers' conceptualizations of creativity: A review and synthesis of contemporary research. *Educational Research Review*, 5, 210-219.
- Asimov, M. (1962). *Introduction to design*. Englewood Cliffs, N.J.: Prentice-Hall.

B

- Baer, J., & McKool, S. (2009). Assessing creativity using the consensual assessment. In C. Schreiner (Ed.), *Handbook of assessment technologies, methods, and applications in higher education*. Hershey, Pennsylvania: IGI Global.
- Baer, J., Kaufman, J. C., & Gentile, C. A. (2004). Extension of the consensual assessment technique to nonparallel creative products. *Creativity Research Journal*, 16, 113–117.
- Basadur, M., Taggar, S. & Pringle, P. (1999). Improving the measurement of divergent thinking. *The Journal of Creative Behavior*, 33, 75-111.
- Beghetto, R. A., & Kaufman, J. C. (2007). Toward a broader conception of creativity: A case for “mini-c” creativity. *Psychology of Aesthetics, Creativity, and the Arts*, 1, 13-79.
- Beghetto, R. A., & Kaufman, J. C. (2009). Intellectual estuaries: Connecting learning and creativity in programs of advanced academics. *Journal of Advanced Academics*, 20, 296-324.
- Besemer, S. P., & O'Quin, K. (1987). Creative product analysis: Testing a model by developing a judging instrument. In S. G. Isaksen (Ed.), (pp. 367-389) *Frontiers of creativity research: Beyond the basics*. Buffalo, NY: Bearly.
- Besemer, S. P. (1998). Creative product analysis matrix: testing the model structure and a comparison among products – three novel chairs. *Creativity Research Journal*, 11, 333–346.
- Besemer, S. P., & O'Quin, K. (1999). Confirming the three-factor creative product analysis model in an American sample. *Creativity Research Journal*, 12, 287–296.
- Birenbaum, M. (2007). Evaluating the assessment: Sources of evidence for quality assurance. *Studies in Educational Evaluation*, 33, 29-49.
- Bloom, B. (1985). *Developing talent in young people*. New York: Ballantine.
- Boden, M. (1990). *The creative mind: Myths and mechanisms*. London : Weidenfeld & Nicolson.
- Bonnardel, N., & Marmèche, E. (2001). Creative design activities : the evocation process and its evolution with regard to expertise. In J.S. Gero & M.L. Maher (Eds.), *Computational and Cognitive Models of Creative Design V* (pp. 189-204). University of Sydney: Key Centre of Design Computing and Cognition, Australia.
- Bonnardel, N. (1989). *L'évaluation de solutions dans la résolution de problèmes de conception*. Rapport de recherche n° 1072. Rocquencourt : INRIA.
- Bonnardel, N. (1991). Criteria used for the evaluation of design solutions. In Y. Queinnec & F. Daniellou (Eds.). *Designing for Everyone and Everybody* (pp. 1043-1045). London: Taylor & Francis.

- Bonnardel, N., Didierjean, A., & Marmèche, E. (2003). Analogie et résolution de problèmes. In C. Tijus (Ed.), *Méthaphores et Analogies* (pp. 115-149). Paris : Hermès.
- Bonnardel, B., Lanzone, L. & Sumner, S. (2003). Designing web sites: opportunistic actions and cognitive effort of lay-designers. *Cognitive Science Quarterly*, 3, 25-56.
- Bonnardel, N. (1993). Comparison of evaluation processes in design activities and critiquing systems: A way to improve design support systems. (Technical Report CU-CS-681-93). Boulder: University of Colorado.
- Bonnardel, N. (1999). L'évaluation réflexive dans la dynamique de l'activité du concepteur. In J. Perrin (Ed.), *Pilotage et évaluation des activités de conception* (pp. 87-105). Paris: L'Harmattan.
- Bonnardel, N. (2000). Towards understanding and supporting creativity in design: Analogies in a constrained cognitive environment. *Knowledge-Based Systems*, 13, 505-513.
- Bonnardel, N. (2006). *Créativité et conception approches cognitives et ergonomiques*. Marseille : Solal.
- Bonnardel, N. (2009). Activités de conception et créativité : De l'analyse des facteurs cognitifs à l'assistance aux activités de conception créatives. *Le Travail Humain*, 72, 5-22.
- Bonnardel, N. (2012). Designing future products: What difficulties do designers encounter and how can their creative process be supported? *Work*, 41, 5296-5303.
- Bonnardel, N., Mazon, S., Wojtczuk, A. (2013). Impact of project-oriented educational methods on creative design *Proceedings of the 31st annual Conference of the European Association of Cognitive - ECCE 2013*. Toulouse, France.
- Bonnardel, N., & Sumner, T. (1996). Supporting evaluation in design. *Acta Psychologica*, 91, 221-244.
- Bonnardel, N., & Zenasni, F. (2010). The impact of technology on creativity in design: An enhancement? *Creativity & Innovation Management*, 19(2), 180-191.
- Bonnardel, N., & Marmèche, E. (2004). Evocation process by Novice and Expert Designers : Towards Stimulating Analogical Thinking. *Creativity and Innovation Management*, 13, 176-186.,
- Bouchard C, Aoussat A. (2003). Design process perceived as an information process to enhance the introduction of new tools. *International Journal of Vehicle Design*, 31, 162-175.
- Bouchard, C., Mantelet, F., Aoussat, A., Selves, C., Gonzales, J.C., Pearce, K., & Coleman, S. (2009). A European emotional investigation in the field of shoes design. *International Journal of Product Development*, 7, 3-27.
- Brangier, E., & Bornet, C. (2011). Persona: A method to produce representations focused on consumers' needs. In W. Karwowski, M. Soares & N. Stanton (Eds). *Human Factors and ergonomics in Consumer Product Design*. Taylor and Francis, pp38-61.

Brunswik, E. (1955). Representative design and probabilistic theory in a functional psychology. *Psychological review*, 62(3), 193-217.

C

Caroff, X., & Besançon, M. (2008). Variability of creativity judgments. *Learning and Individual Differences*, 18(4), 367-371

Chan D.W., & Chan, L. (1999). Implicit theories of creativity: teachers' perception of students characteristics in Hong-Kong. *Creativity Research Journal*, 12, 185-195.

Chen, C., Kasof, J., Himsel, A., Greenberger, E., Dong, Q., & Xue, G. (2002). Creativity in drawings of geometric shapes: A cross-cultural examination with the Consensual Assessment Technique. *Journal of Cross-Cultural Psychology*, 33, 171-187.

Christiaans, H., & Venselaar, K. (2005). Creativity in Design Engineering and the Role of Knowledge: Modeling the Expert. *International Journal of Technology and Design Education*, 15, 217-236.

Christiaans, H. (2002). Creativity as a design criterion. *Creativity Research Journal*, 14, 41-54.

Conti, R., & Amabile, T.M. (1995). *Problem solving among computer science students: The effects of skill, evaluation, expectation and personality on solution quality*. Paper presented at the annual meeting of the Eastern Psychological Association, Boston, Mass.

Craft, A. (2005). *Creativity in schools: tensions and dilemmas*. Abingdon: Routledge.

Cropley, D. H., & Cropley, A. J. (2010). Functional Creativity: Products and the Generation of Effective Novelty. In J. C. Kaufman and R. J. Sternberg (Eds.), (pp. 301-317). *Cambridge Handbook of Creativity*. New York: Cambridge University Press.

Cropley, D.H. & Cropley, A. J. (2005). Engineering creativity: A systems concept of functional creativity. In J. C. Kaufman and J. Baer (Eds.), (pp. 169-185). *Creativity across domains: Faces of the muse*. New Jersey: Lawrence Erlbaum.

Cropley, D. H., Kaufman, J. C., & Cropley, A. J. (2011). Measuring Creativity for Innovation Management. *Journal of Technology Management & Innovation*, 6, 13-30.

Cross, N. (1997). Descriptive models of creative design: Application to an example. *Design Studies*, 18, 427-455.

Cross, N. (2004). Expertise in design: An overview. *Design Studies*, 25, 427-441. Csikszentmihalyi, 1988;

Csikszentmihalyi, M. (1999). Implication of a system perspective for the study of creativity. In R.J. Sternberg (Ed.), *Handbook of creativity* (pp. 313-339). New-York: Cambridge University Press

D

Dawkins, R. (1976). *The selfish gene*. New York City: Oxford University Press.

Decision No 1350/2008/EC of the European Parliament and of the Council of 16 December 2008 concerning the European Year of Creativity and Innovation (2009). Official Journal of the European Union.

Demirkan, H., & Afacan, Y. (2012). Assessing creativity in design education: Analysis of creativity factors in the first-year design studio. *Design Studies*, 33(3), 262-278.

Demirkan, H., & Hasirci, D. (2009). Hidden Dimensions of Creativity Elements in Design Process. *Creativity Research Journal*, 21(2-3), 294-301.

Do, E. Y.-L., Gross, M. D., Neiman, B., & Zimring, C. (2000). Intentions in and relations among design drawings. *Design Studies*, 21(5), 483-503.

Dollinger, S.J. (2007). Creativity and conservatism. *Personality and Individual Differences*, 43, 1025-1035.

Dollinger, S. J., & Shaffran, M. (2005). Note on consensual assessment technique in creativity research. *Perceptual and Motor Skills*, 100, 592-598.

Dollinger, S. J., Urban, K. K., & James, T. A. (2004). Creativity and openness: Further validation of two creative product measures. *Creativity Research Journal*, 16, 35-47.

Dorst, K., & Dijkhuis, J. (1995). Comparing paradigms for describing design activity. *Design Studies*, 16, 261-274.

Dorst, K., (1997). *Describing Design - A Comparison of Paradigms*, Thesis TUDelft, The Netherlands.

Dorst, K., & Cross, N. (2001). Creativity in the design process: Co-evolution of problem-solution. *Design Studies*, 22, 425-437.

E

Eastman, C.M. (1969). Cognitive processes and ill-defined problems: a case study from design. *Proceedings of the 1st International Joint Conference on I.A.* (pp. 669-690). Washington, D.C.

F

- Feist, G.J. (1998). A meta-analysis of personality in scientific and artistic creativity. *Personality & Social Psychology Review (Lawrence Erlbaum Associates)*, 2, 209-309.
- Fenton, R. (1996). Performance assessment system development. *Alaska Educational Research Journal*, 2, 13-22.
- Finke, R. A., Ward, T. B., & Smith, S. M. (1992). *Creative cognition : theory, research, and applications*. Cambridge, Mass.: MIT Press.
- Fustier, M. (1989). *La résolution de problème : méthodologie de l'action*. Paris : Editions ESF & Librairies Techniques.

G

- Gardner, H. (1993). *Multiple intelligences : the theory in practice*. New York: Basic Books.
- Gelb, M. (1996). *Thinking for a change: discovering the power to create, communicate and lead*. New York: Crown Books.
- Gero, J.S., & Kannengiesser, U. (2004). The situated Function-Behaviour-Structure framework. *Design Studies*, 25, 373-391.
- Gero, J.S. (1990). Design prototypes: a knowledge representation scheme for design. *AI Magazine*, 11, 26-36.
- Gero, J. S. (2000). Computational models of innovative and creative design processes. *Technological Forecasting and Social Change*, 64, 183-196.
- Gigerenzer, G. (2008). *Rationality for mortals: How people cope with uncertainty*. New York: Oxford University Press.
- Glaser, B.G. & Strauss, A.L. (1967). *The discovery of grounded theory: Strategies for qualitative research*. New York: Aldine de Gruyter
- Glăveanu, V.P. (2010). Creativity in context: The ecology of creativity evaluations and practices in an artistic craft. *Psychology Studies*, 55, 339-350.
- Glăveanu, V.P. (2012). A multiple feedback methodology for the study of creativity evaluations. *Journal of Constructivist Psychology*, 25, 346-366.
- Goel, V. (1995). *Sketches of Thought*. Cambridge: MIT Press.
- Goldschmidt, G. (2001). Is a figure-concept binary argumentation inherent in visual design reasoning? In J.S. Gero, B. Tversky, & T. Purcell (Eds.), *Visual and spatial reasoning in design II* (pp. 177-206). Sydney: University of Sydney.
- Goulding, C. (2002). *Grounded Theory; A Practical Guide for Management, Business and Market Researchers*. London: Sage Publications.

Gross, M & Do, E. (1996). Ambiguous Intentions : A paper-like Interface For Creative Design. *ACM (ed), Proceedings of ACM Conference on User Interface Software Technology (UIST 96)* (pp. 183-192). Seattle.

Guilford, J.P. (1967). *The nature of human intelligence*. New York : McGraw Hill.

Guilford, J.P. (1981). Potentiality for creativity. In J.C. Gowan, J. Khatena, & E.P. Torrance (Eds.), *Creativity: Its educational implications* (pp. 1-5). Dubuque, IA: Kendall Hunt.

H

Hasirci, D., & Demirkan, H. (2003). Creativity in learning environments: The case of two sixth grade art-rooms. *Journal of Creative Behavior*, 37, 17–42.

Hayes-Roth, B. & Hayes-Roth, F. (1979). A cognitive model of planning. *Cognitive Science*, 3, 275-310.

Hennessey, B. A. (2003). The Social Psychology of Creativity. *Scandinavian Journal of Educational Research*, 47(3), 253-271.

Hickey, M. (2001). An application of Amabile's consensual assessment technique for rating the creativity of children's musical compositions. *Journal of Research in Music Education*, 49, 234-244.

C. Hooker, J. Nakamura, & M. Csikszentmihalyi. The group as mentor: Social capital and the systems model of creativity. In P. Paulus and M. Nijstad (Eds.), *Group Creativity*, pp. 225-44. (2003)

Horn, D. & Salvendy, G. (2006). Consumer-based assessment of product creativity: A review and reappraisal. *Human Factors and Ergonomics in Manufacturing*, 16, 155-176.

Horn, D., & Salvendy, G. (2006). Product creativity: conceptual model, measurement and characteristics. *Theoretical Issues in Ergonomics Science*, 7, 395-412.

Horn, D., & Salvendy, G. (2009). Measuring consumer perception of product creativity: impact on satisfaction and purchasability. *Human Factors and Ergonomics in Manufacturing*, 19, 223-240.

Hsiao, S.W., Chan, J.R. (2004). A creativity-based design process for innovative product design. *International Journal of Industrial Ergonomics*, 34, 421-443.

J

Jackson, P. & Messick, S. (1965). The person, the product and the response: Conceptual problems in the assessment of creativity. *Journal of Personality*, 33, 309-239.

- Jeffries, K. K. (2012). Amabile's consensual assessment technique: Why has it not been used more in design creativity research? *Proceedings of the 2nd International Conference on Design Creativity* (pp. 211-220). Glasgow.
- Johnson, H. & Carruthers, L. (2006). Supporting creative and reflective process. *International Journal of Human -Computer Studies*, 64, 998-1030.
- Johnson-Laird, P.N. (1995). Mental models, deductive reasoning, and the brain, and the brain. In M.-S. Gazzaniga (Ed.), *The cognitive neurosciences* (pp. 999-1008). Cambridge: M.I.T. Press.
- John-Steiner, V. (1992). Creative lives, creative tensions. *Creativity Research Journal*, 5(1), 99-108.

K

- Kaufman, J. C. (2006). Self-reported differences in creativity by ethnicity and gender. *Applied Cognitive Psychology*, 20, 1065–1082.
- Kaufman, J. C., Baer, J., Cole, J. C., & Sexton, J. D. (2008). A comparison of expert and nonexpert raters using the consensual assessment technique. *Creativity Research Journal*, 20, 171-178.
- Kaufman, J. C., Baer, J., & Cole, J. C. (2009). Expertise, domains, and the Consensual Assessment Technique. *Journal of Creative Behavior*, 43, 223–233.
- Kaufman, J. C., Gentile, C. A., & Baer, J. (2005). Do gifted student writers and creative writing experts rate creativity the same way? *Gifted Child Quarterly*, 49, 260–265.
- Kirkwood, B.R., & Sterne, J.A.C. (2003). *Essential medical statistics*. Malden, MA: Blackwell Science
- Kosslyn, S.M. (1980). *Image and mind*. Cambridge: Harvard University Press.
- Kreitler, S., & Casakin, H., (2009). Self-perceived creativity: the perspective of design. *European Journal of Psychological Assessment*, 25(3), 194-203.

L

- Lawson, B., & Dorst, K. (2009). *Design Expertise*. Oxford: Architectural Press.
- Lebahar, J.C. (1983). *Le dessin d'architecte*. Roquevaire: Parenthèse.
- Leder, H., Augustin, D., & Belke, B. (2005). Art and cognition! Consequences for experimental aesthetics. *Bulletin of Psychology and the Arts*, 5, 11-20.
- Lee, S, Lee J. and Young, C-Y. (2005). A variation of CAT for measuring creativity in business products. *Korean Journal of Thinking and Problem Solving*, 15, 143-153.
- Leplat, J. (1985). Les représentations fonctionnelles dans le travail. *Psychologie Française*, 30, 269-275.

- Leplat, J. (1997). *Regards sur l'activité en situation de travail : Contribution à la psychologie ergonomique*. Paris : PUF.
- Leplat, J., & Pailhous, J. (1981). L'acquisition des habiletés mentales, la place des techniques. *Le travail Humain*, 44, 275-284.
- Lera, S. (1981). Empirical and theoretical studies of design judgment: a review. *Design Studies*, 1, 19-26.
- Lera, S. (1981). Architectural designers' values and the evaluation of their designs. *Design Studies*, 2, 131-137
- Lubart, T.I. (1999). Componential models of creativity. In M.A. Runco & S. Pritzer (Eds.), *Encyclopedia of creativity* (pp. 295-308). New York: Academic Press.
- Lubart, T.I., Mouchiroud, C., Tordjman, S. & Zenasni, F. (2003). *Psychologie de la créativité*. Paris: Armand Colin.
- Lubart, T. (1994). Creativity. In R.J. Sternberg (Ed.), *Thinking and Problem Solving* (pp. 289-332). New York: Academic Press.
- Lubart, T. (2005). How can computers be partners in the creative process. *International Journal of Human-Computer Studies*, 63, 365-369.
- Lynch, P. and S. Horton. 2008. *Web Style Guide: Basic Design Principles for Creating Web Sites*. New Haven: Yale University Press.

M

- Martindale, C. (1990). *The clockwork muse: the predictability of artistic change*. New York: Basic Books.
- Mayer, R. E. (1989). Models for understanding. *Review of Educational Research*, 59, 43-64.
- Mayer, R. E. (1999). Fifty years of creativity research. In R.J. Sternberg (Ed.), *Handbook of creativity* (pp. 449-461). New-York: Cambridge University Press.
- McNeill, T., Gero, J., & Warren, J. (1998). Understanding conceptual electronic design using protocol analysis. *Research in Engineering Design*, 10(3), 129-140.
- Milton, F. (2003). Innovation survey. last accessed on 04/20/2014 at http://www.wowgreatidea.com/articles/7_PWCInnovationSurvey.pdf.
- Mouchiroud, C., & Bernoussi, A. (2008). An empirical study of the construct validity of social creativity. *Learning and Individual Differences*, 18, 372-380.

N

- Newell, A., & Simon, H. A. (1972). *Human problem solving*. Englewood Cliffs, N. J.: Prentice-Hall.
- Norman, D.A. (1988). *The Design of Everyday Things*. London: MIT Press.
- Norman, D.A. (2004). *Emotional Design: Why We Love (or hate) Everyday Things*. New York: Basic Books.

P

- Pektas, S.T. (2010). Effects of cognitive styles on 2D drafting and design performance in digital media. *International Journal of Technology and Design Education*, 20, 63-76.
- Piffer, D. (2012). Can creativity be measured? An attempt to clarify the notion of creativity and general directions for future research. *Thinking Skills and Creativity*, 7, 258-264.
- Plucker, J. A., Beghetto, R. A., & Dow, G. T. (2004). Why isn't creativity more important to educational psychologists? Potentials, pitfalls, and future directions in creativity research. *Educational Psychologist*, 39, 83-96.
- Plucker, J. A., Holden, J., & Neustadter, D. (2008). The criterion problem and creativity in film: Psychometric characteristics of various measures. *Psychology of Aesthetics, Creativity, and the Arts*, 2, 190-196.
- Plucker, J. A., & Makel, M. C. (2010). Assessment of creativity. In J. C. Kaufman & R. J. Sternberg (Eds.), *The Cambridge Handbook of Creativity* (pp. 48-73). Cambridge: Cambridge University Press.
- Plucker, J. A., Kaufman, J. C., Temple, J. S., & Qian, M. (2009). Do experts and novices evaluate movies the same way? *Psychology and Marketing*, 26, 470-478.
- Policastro, E. and Gardner, H. (1999). From case studies to robust generalizations, an approach to the study of creativity. In R.J. Sternberg (Ed.), *Handbook of Creativity* (pp. 213-225). New York: Cambridge University Press.
- Priest, T. (2006). Self-evaluation, creativity, and musical achievement. *Psychology of Music*, 34, 47-61.
- Pruitt, J. & Adlin, T. (2006). *The Persona Lifecycle*. Morgan Kaufmann, San Francisco, CA.

R

- Reis, S. M., & Renzulli, J. S. (1991). The assessment of creative products in programs for gifted and talented students. *Gifted Child Quarterly*, 35, 128-134
- Reitman, W.R. (1964). Heuristic decision procedures, open constraints, and the structure of ill-defined problems. In M.W. Shelly & G.L. Bryan (Eds.), *Human judgements and optimality*. New York: Wiley.

- Rhodes, M. (1961). Analysis of Creativity. *Phi Delta Kappan* 42, 305–310.
- Richard, J.-F. (1995). *Les activités mentales : comprendre, raisonner, trouver des solutions*. Paris : Armand Colin.
- Rittel, H. & Webber, M.M. (1984). Planning problems are wicked problems. In N. Cross (Ed.), *Developments in Design Methodology* (pp. 135-144). New York: John Wiley & Sons, Inc.
- Runco, M.A. (2004). Creativity. *Annual Review of Psychology*, 55, 657-687.
- Runco, M.A. and Basadur, M.S. (1993). Assessing ideational and evaluative skills and creative styles and attitudes. *Creativity and Innovation Management*, 2, 166-173.
- Runco, M.A., & Charles, R. (1993). Judgments of originality and appropriateness as predictors of creativity. *Personality and Individual Differences*, 15, 537-546.
- Ruscio, J., Whitney, D., & Amabile, T.M. (1995). How do motivation and task behaviors affect creativity? An investigation in three domains. Unpublished manuscript, Brandeis University.

S

- Sarkar, P., & Chakrabarti, A. (2008). Studying engineering design creativity – developing a common definition and associated measures. In J. Gero (Ed.), *Studying design creativity*. Springer Verlag .
- Schaefer, C. E. (1969a). The prediction of creative achievement from a biographical inventory. *Educational and Psychological Measurement*, 29, 431–37.
- Schneiderman, B. (2000). Creating creativity: user interfaces for supporting innovation. *ACM Transactions on Computer-Human Interaction*, 7, 114-138.
- Schön, D.A. & Wiggins, G. (1992) Kinds of seeing and their functions in designing. *Design Studies*, 13, 135-156.
- Schön, D. A. (1988). Designing: Rules, types and worlds. *Design Studies*, 9, 181-190.
- Schön, D.A. (1983). *The reflective practitioner: How professionals think in action*. New York: Basic Books.
- Shah, J., Vargas-Hernandez, N., Smith, S. (2003). Metrics for measuring ideation effectiveness. *Design Studies*, 24, 111-143.
- Simon, H. A. (1956). Rational choice and the structure of the environment. *Psychological Review* 63, 129-38.
- Simon, H.A. (1995). Problem forming, problem finding and problem solving in design. In A. Collen & W Gasparski (Eds.), *Design & Systems* (pp. 245-257). New Brunswick: Transaction Publishers.
- Simon, H.A. (1969). *The sciences of the artificial*. Cambridge: MIT Press.
- Simon, H.A. (1973). The structure of ill structured problems. *Artificial Intelligence*, 4, 181-201.

- Simonton, D. K. (2000). Creative development as acquired expertise: theoretical issues and an empirical test. *Developmental Review*, 20, 283-318.
- Simonton, D. K. (2010). Creative thought as blind-variation and selective-retention: Combinatorial models of exceptional creativity. *Physics of Life Reviews*, 7(2), 156-179.
- Simonton, D. K. (1995). *Creative expertise: A life-span developmental perspective*. Paper presented at the Acquisition of expert performance: An international conference., Wakulla Springs.
- Smith, E. E., Jonides, J., Koeppel, R. A., Awh, E., Schumacher, E. H., & Minoshima, S. (1995). Spatial vs. object working memory: PET investigations. *Journal of Cognitive Neuroscience*, 7(337-356).
- Sosa, R., & Gero, J.S. (2004). A computational framework for the study of creativity and innovation in design: Effects of social ties. In J.S. Gero (Ed.) *Design Computing and Cognition '04* (pp. 499-517), Dordrecht: Kluwer Academic Publishers.
- Stacey, M., & Eckert, C. (2003). Against ambiguity. *Computer Supported Cooperative Work*, 12, 153–183.
- Sternberg, R.J. & Lubart, T.I. (1995). *Defying the crowd: creativity in a culture of conformity*. New York: Free Press.
- Sternberg, R.J. (1999). *Handbook of creativity*. New York: Cambridge University Press.
- Sternberg, R.J., Kaufman, J.C. & Pretz, J.E. (2002). The creativity conundrum: a propulsion model of kinds of creative contribution. New York: Psychology Press.
- Sternberg, R. J., Kaufman, J. C., & Pretz, J. E. (2003). A propulsion model of creative leadership. *Leadership Quarterly*, 14, 455-473.
- Sternberg, R.J., & Lubart, T. (1999). The concept of creativity: Prospects and paradigms. In R.J. Sternberg (Ed.), *Handbook of creativity* (pp. 3-15). New-York: Cambridge University Press.

T

- Taylor, C. W., & Ellison, R. L. (1966). *Alpha biographical inventory*. Salt Lake City: Institute for Behavioral Research in Creativity.
- Taylor, C. W., & Ellison, R. L. (1967). Biographical predictors of scientific performance. *Science* 155, 1075–1079.
- Taylor, A. (1975). An emerging view of creative actions. In I. A. Tylor, & J. W. Getzels (Eds.), *Perspectives in creativity* (pp. 297-325). Chicago: Aldine.
- Torrance, E. P., & Khatena, J. (1970). What Kind of Person Are You?: a brief screening device for identifying creatively gifted adolescents and adults. *Gifted Child Quarterly*, 14, 71-75.
- Torrance, E. P. (1962). *Guiding creative talent*. Englewood Cliffs: Prentice Hall.

- Torrance, E. P. (1974). *Torrance Tests of Creative Thinking Norms-Technical Manual*. Lexington, Mass: Ginn and Co.
- Tractinsky, N., Shoval-Katz, A., & Ikar, D. (2000). What is Beautiful is Usable. *Interacting with Computers*, 13, 127–145.
- Tversky, A. & Kahneman, D. (1974). Judgment under uncertainty: Heuristics and biases. *Science*, 185, 1124-1131.
- Tversky, B. (2003). Structures of mental spaces: How people think about space. *Environment and Behavior*, 35, 66-80.

V

- Van der Lugt, R. (2000). Developing a graphic tool for creative problem solving in design groups. *Design Studies*, 21, 505-522.
- Visser, W. (2002). Conception individuelle et collective. Approche de l'ergonomie. In M. Borillo & J.-P. Goulette (Eds.), *Cognition et création. Explorations cognitives des processus de conception* (pp. 311-327). Bruxelles: Mardaga.
- Visser, W. (2006). *The cognitive artifacts of designing*. Mahwah: Lawrence Erlbaum Associates.
- Visser, W. (2010). Visser : design as construction of representations. *Collection, Art + Design & Psychology*, 2, 29-45.
- Visser, W. (2009a). La conception: De la resolution de problems à la construction de représentations. *Le travail Humain*, 72(1), 61-78.
- Visser, W., Darses, F., & Détienne, F. (2004). Approches théoriques pour une ergonomie cognitive de la conception. In J-M Hoc & F. Darses (Eds.), *Psychologie ergonomique, tendances actuelles* (pp. 97-118). Paris : Presse Universitaire de France.

W

- Wallach, M. A., & Kogan, N. (1965). *Modes of thinking in young children: A study of the creativity intelligence distinction*. New York: Holt, Rinehart & Winston.
- Wallas, G. (1926). *The art of thought*. New York: Harcourt-Brace.
- Ward, T.B., Smith, S.M., & Finke, R.A. (1999). Creative cognition. In R.J. Sternberg (Ed.), *Handbook of creativity* (pp. 189-212). New-York: Cambridge University Press.

- West, M. A. (2002). Sparkling fountains or stagnant ponds: An integrative model of creativity and innovation implementation in work groups. *Applied Psychology: An International Review*, 51, 355-424.
- Williams, W. M. and Yang, L.T. (1999). Organizational creativity. In R.J. Sternberg (Ed.), *Handbook of creativity* (pp. 373-391). New-York: Cambridge University Press.
- Wojtczuk, A., & Bonnardel, N. (2010). L'effet de l'outil sur la qualité des productions. In C. Yacoub & K. Zreik (Eds.), *Conception assistée par concepteur* (pp. 53-60). Paris, France : Europia Productions.
- Wojtczuk, A., & Bonnardel, N. (2010). Externalisation in design: impact of different tools on designers' activities and on the assessment of final design. In W. Brinkman & M. Neerincx (Eds.) *Proceedings of the 28th annual Conference of the European Association of Cognitive - ECCE 2010* (pp. 75-80). Delft, The Netherlands: Mediamatrica
- Wojtczuk, A., & Bonnardel, N. (2011). Designing and assessing everyday objects: Impact of externalisation tools and judges' backgrounds. *Interacting with Computers*, 23(4), 337-345.

Y

- Yan, H., Huynh, V., Murai, T., & Nakamori, Y. (2008). Kansei evaluation based on prioritized multi-attribute fuzzy target-oriented decision analysis. *Information Sciences*, 178, 4080–4093.

Z

- Zeisel, J. (1981). Inquiry by design, tools for environmental behavior research. Cambridge: Cambridge University Press.
- Zeng, L. Salvendy, G. & Zhang, M. (2009). Factor structure of web site creativity. *Computers in Human Behavior*, 25, 568-577.

APPENDICES

Appendix 1 - Online questionnaire used for the first study

Critères d'évaluation de la créativité

Bonjour, Je m'appelle Alicja Wojteczuk, je réalise une thèse à l'Université d'Aix-Marseille. Le sujet de ma recherche porte sur le processus d'évaluation de la créativité en design. Pour réaliser ce travail, je sollicite des personnes ayant une expérience d'au moins 5 ans dans le domaine du design. Si vous remplissez ce critère, merci de participer à ma recherche en répondant aux quelques questions ci-dessous. Ce questionnaire ne vous prendra que 5 minutes. Merci d'avance pour vos réponses. Alicja Wojteczuk

* Required

Vous êtes : *

- ☒ Homme
- ☐ Femme

e-Mail : *Votre adresse mail ne sera en aucun cas utilisé à des fins commerciales.

Connaitre votre expertise métier

Quelle est votre profession ? *Vous pouvez sélectionner plusieurs professions

- ☐ A - Designer graphique
- ☐ B - Designer d'objets
- ☐ C - Directeur Artistique
- ☐ D - Enseignant en Design
- ☐ E - Commerçant dans le domaine du Design

- ☐ Other:

Depuis combien d'années exercez-vous cette profession ? *

Si plusieurs professions, indiquez un nombre d'années pour chacune dans l'ordre (A - xx ; B - xx ...)

Questions ouvertes, vous pouvez y répondre librement

Pour les 2 questions suivantes, vous pouvez citer autant de critères (mot ou expression) que vous le voulez, en étant le plus explicite possible.

Notez les critères qui vous permettent d'évaluer la créativité en design GRAPHIQUE : *Merci de séparer chacun d'eux par un ;

Notez les critères qui vous permettent d'évaluer la créativité en design d'OBJET : *Merci de séparer chacun d'eux par un ;

Allons plus loin ensemble

Seriez-vous susceptible de participer à la seconde partie de ma recherche (sous forme d'un autre questionnaire) ? *

- ☒ Oui
- ☐ Non

Submit

Appendix 2 - Evaluative referents for Graphic Design (GD) and Product Design (PD)

Table 33: Categorized ERs mentioned as the most important for creativity in both, graphic and product design areas (during the first study).

			Examples of used ER	
Criterion		Definition	GD	PD
Originality	GD PD	reference to the innovation, uncommonness	<i>Originalité; Présenter le message de manière originale; Originalité du traitement; Nouveau; Innovation; Surprenant; Rupture ; Insolite; Différence; Sortant de l'ordinaire; Amener une idée nouvelle</i>	<i>Originalité; Surprise; Nouveauté; Innovation; Inconnu; Exploration peu couvert par design actuel; Rupture ; Différence ; Apport technique innovant; Apport dans l'usage</i>
Creative-person qualities	GD PD	Describes the characteristics considered as important for creative person	<i>Inventivité ; Imagination; Ingéniosité; Talent; Audace; Culot; Parti pris</i>	<i>Ingéniosité; Inventivité Intelligence; Réfléchi; Audace; Anticonformisme; Culot; Décalé; Capacité à communiquer ses idées</i>
Creative-process qualities	GD PD	Divergent thinking and visibility of the creative process within the final production	<i>Multiplicité des réponses; Idées plurielles; Variété; Directions de réponses différentes ; Evolution dans la démarche; Reflète la démarche de la problématique</i>	<i>Pistes de recherche multiples; Multiplicité des réponses; Variété</i>
Concept	GD PD	Quality and the elaboration of the idea on which the design is based	<i>Concept; Conceptualisation; Idée</i>	<i>Concept; Idée</i>
Relevance with the subject	GD	importance of the semantic connection between the content and the theme of the graphic design	<i>Respect de cible; Pertinence par rapport à la demande; Pertinence; Pertinence des symboles; Symboles et icônes judicieusement choisis; Rapport avec la thématique; Evocation des couleurs; Sens de l'image; Sens ; Crédibilité; Cohérent; Justesse; Message; Idées évoquées</i>	
Comprehension of the message	GD	Degree to which the idea represented in the	<i>Compréhension; Clarté du message; Clair; Lisibilité; Accessibilité; Représentation précise ;</i>	

		design is easy to understand for the public	<i>Forme qui ne parasite pas la lecture</i>	
Comprehension of the use	PD	degree to which it is comprehensive for the user to understand the design functionalities		<i>Sens véhiculé ; Respect de cahier des charges; Pertinence par rapport à une cible; Appropriation par usager; Adaptabilité; Adapté; Pertinence ; Justesse; Affordant; Prise en main facilitée par le design; Lisibilité; Représenter la fonction ; Fiabilité; Analogie; Associations pertinentes; Imitation</i>
Ergonomics and user-appropriateness	GD PD	connected with usability of the design	<i>Ergonomie; Adapté; Facile à utiliser; Fonctionnalité; Réponse au besoin; Efficacité; Utilité</i>	<i>Ergonomie; Praticité; Pratique; Utilisabilité; Confortable; Interaction; Maniabilité; Manipulation; Fonctionnalité ; Efficace; Efficience; Fonctionnalités ; Fonction; Utilisé face à un manque; Réponse à un besoin d'utilisateur; Scénario d'usage; Correspondance au Contexte d'usage; Utilité</i>
Respecting the codes of the domain	GD PD	field codes, existing trends and other elements facilitating the recognition	<i>Modernité; Dans l'air du temps; Connaissance du milieu; Tendance; Références; Faire référence au passé; Cohérence avec le monde actuel; Déjà vu; Codes culturels; Imitation; Rapprochement par rapport à ce que je connais déjà; Culture ; Références à d'autres à d'autres cultures; Culture esthétique; Associations; Universalité</i>	<i>Modernité; Actuel; Réinvestissement des références; Connaissance du milieu; Cohérence avec le Monde actuel; Histoire de l'art; Appel à la mémoire Collective; Culture esthétique; Intemporel; Culture</i>
Aesthetics	GD PD	the appearance of design	<i>Esthétique; Beauté</i>	<i>Aspect; Esthétique; Beau; Attirance esthétique; Laisser une impression du "beau" ou "mieux"; Artistique</i>
Design elements	GD PD	visual components of design	<i>Couleur ; Typographie; Image; Format; Forme; Contrastes</i>	<i>Forme; Ligne de l'objet; Couleur; Taille; Volume; Perception de l'espace, Perspective; Matériaux;</i>

				<i>Utilisation de nouvelles Matières; Lumière; Matière; Texture; Toucher</i>
Practical constraints	PD	constraints related with life of product, production process and marketing domains important to take into account within a design product		<i>Gestion des contraintes; Apport de nouvelles contraintes; Gain de place; Economie; Durabilité; Rapport coût de prod. - qualité - prix; Respect de marketing; Packaging; Emergence par rapport au secteur; Marché</i>
Values	PD	moral constraints to be taken into account within a design product		<i>Responsabilité ; Ethique; Equitable; Ecologie</i>
Harmony	GD PD	the way in which the design elements should be adjusted with each other to create an impression of unity	<i>Homogénéité; Unicité; Harmonie; Harmonie des formes; Harmonie couleur; Harmonie typographique; Equilibre de composition, grille, rapports de taille, utilisation des blancs; Equilibre ; Equilibre des volumes; Rapport fond / forme; Cohérence entre contenu et forme</i>	<i>Equilibre des formes; Harmonie des formes; Equilibre des volumes; équilibre; Harmonie; Harmonie des couleurs; Respiration; Rythme; Unicité; Couleurs et formes cohérentes</i>
Layout	GD	organization of the design elements on the given surface	<i>Utilisation de l'espace; Mise en page; Respiration; Rythme; Composition; Organisation; Ingéniosité du layout; Hiérarchisation; Disposition réfléchie des éléments</i>	
Quality of execution	GD PD	precision and finishing of the design work	<i>Précision; Qualité; Qualité plastique; Qualité de réalisation; Qualité de mise en œuvre; Qualité technique; Qualité d'image; Exécution; Détail; Rigueur de production; Finition; Finesse</i>	<i>Finition; Qualité; Précision; Qualité de fabrication; Qualité des matériaux; Détails qui donnent la qualité</i>
Style	GD PD	underlines the importance of some strong, recognizable traits	<i>Personnel; Ecriture personnelle; Identitaire; Dégagement d'un style; Style; Style de dessin</i>	<i>Personnalité; Style; Unique</i>
Simplicity	GD PD	Use of simple means and minimum of	<i>Simplicité; Sobriété</i>	<i>Simplicité; Sobriété</i>

		elements		
Emotions conveyed	GD PD	evoking emotional reaction	<i>Emotion; Sensibilité; Expressif; Impression; Ambiance; Plaisir de voir ; Plaisir ; Fun ; Curiosité; Humour; Laisser un souvenir</i>	<i>Valence émotionnelle; Emotion; Sensibilité; Suscite des pensées agréables ou fortes; Plaisir d'utilisation; Plaisir ; Joindre utile à l'agréable; Humour; Amusant; Drôle; Ludique; Curieux; Impression qui s'en dégage</i>
Appeal	GD PD	attracting the target attention	<i>Attractivité; Impact; Impact visuel; Visibilité; Captiver le regard; Interpeler la personne; susciter une envie; Séduisant</i>	<i>Impact; Attractivité; Envie; Exciter la convoitise; Rendre jaloux ; Séduction</i>
Tools	GD PD	technology used during the creation process	<i>Outils utilisées; Technique; Technologie</i>	<i>Choix technologie pour fonction; Technique; Technologie; Prouesse technique</i>

Appendix 3 - Number of evaluative referents cited by judges, depending on their (a) backgrounds and (b) levels of experience in design.

Table 34: ERs cited during the first study, according to the judges backgrounds.

a	designers	art.dir.	teachers	audience	TOTAL
Aesthetics	3	5	1	4	13
Appeal	3	0	2	2	7
Comprehension of the message	6	4	2	6	18
Concept	1	2	0	0	3
Creative-person qualities	0	4	1	3	8
Creative-process characteristics	0	2	2	0	4
Design elements	4	4	1	5	14
Emotions conveyed	5	5	0	5	15
Ergonomics and user-app.	4	4	1	5	14
Harmony	3	5	1	4	13
Layout	3	2	2	2	9
Originality	10	11	7	15	43
Quality of execution	4	3	2	2	11
Relevance with the subject	6	8	6	4	24
Respecting the codes	5	2	1	5	13
Simplicity	4	0	1	2	7
Style	2	4	1	1	8
Tools	1	0	1	1	3

Table 35: ERs cited during the first study, according to the judges levels of experience in design.

b	a. experts	Int. experts	laypeople	TOTAL
Aesthetics	4	5	4	13
Appeal	4	1	2	7
Comprehension of the message	3	9	6	18
Concept	1	2	0	3
Creative-person qualities	2	3	3	8
Creative-process characteristics	3	1	0	4
Design elements	5	4	5	14
Emotions conveyed	3	7	5	15
Ergonomics and user-app.	2	7	5	14
Harmony	6	3	4	13
Layout	2	5	2	9
Originality	16	12	15	43
Quality of execution	4	5	2	11
Relevance with the subject	11	9	4	24
Respecting the codes	2	6	5	13
Simplicity	2	3	2	7
Style	4	3	1	8
Tools	0	2	1	3

Appendix 4 - Indications for participants of third study

Vous allez participer à une étude sur la créativité en design.

Dans ce cadre, je vais vous présenter quelques travaux effectués par les étudiants en design et votre rôle sera de me dire ce que vous pensez de leur niveau créatif. Je vous invite à exprimer tout ce qui vous vient à l'esprit, même s'il y a des choses qui vous paraissent contradictoires ou peu pertinentes. Ce qui m'intéresse, c'est justement votre processus de réflexion.

PHASE 1

- Présentation du cahier des charges : lecture.

Ceci est le cahier des charges présenté aux étudiants en design. Les travaux que vous allez voir ensuite, sont produits par ces étudiants à partir de ce cahier des charges. Merci de le lire avant de passer à l'étape suivante.

- Présentation des affiches dans un ordre aléatoire.

Consigne :

- Regardez ces travaux et choisissez ceux que vous trouvez les plus créatifs en vous basant sur votre première impression.
- Pourquoi avez-vous choisi ces travaux-là ?

PHASE 2

- Présentation des affiches un par un, dans le même ordre que dans la phase 1.

Consigne :

- Maintenant, je vais vous demander de regarder ces travaux encore une fois, une par une. Pour chacun d'eux, je vais vous demander de dire à voix haute ce que vous en pensez et de répondre à quelques questions.
 - Que pensez-vous de la créativité de cette affiche ? /expression libre/
 - Qu'aurez-vous changé dans cette affiche, si vous étiez designer, afin de l'améliorer ?

PHASE 3

- Présentation des affiches dans le même ordre que dans les phases 1 et 2.

Consignes :

- Vous avez regardé attentivement toutes les affiches et vous avez analysé leur créativité. Si vous deviez choisir maintenant, en tenant compte de vos constats durant cette analyse, laquelle/lesquelles de ces affiches considéreriez-vous comme la/les plus créative(s) ?

- Si vous deviez choisir maintenant l’affiche ou les affiches que vous préférez, sans penser à leur créativité, laquelle ou lesquelles sélectionneriez-vous ?
- Pour quelles raisons ?

Appendix 5 - Posters presented to participants of the third study





Appendix 6 - Criteria used during the entire duration of the interviews

- distribution depending on judges' backgrounds

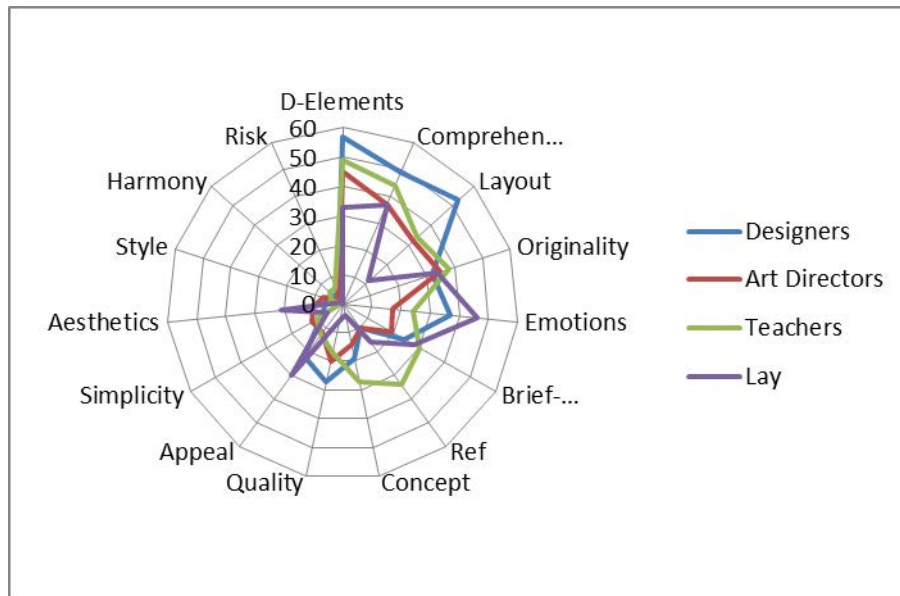


Figure 12: Numbers of occurrences for ERs used during the whole interview, within groups of judges with different backgrounds.

Table 36: Standardized Deviates for the number of occurrences of ERs, within the groups of judges with different professional backgrounds.

	Designers	Art Dir.	Teachers	Lay	
D-Elements	0,56	0,68	-0,11	-1,20	-0,07
Comprehension	0,14	0,02	-0,18	0,02	0,00
Layout	2,50	0,57	-0,25	-3,15	-0,33
Originality	-1,26	0,78	0,06	0,60	0,17
Emotions	0,23	-1,99	-1,65	3,56	0,14
Brief-approp.	-0,94	-0,70	0,51	1,20	0,08
Ref	-2,26	-1,39	3,46	0,14	-0,05
Concept	0,14	-0,04	2,32	-2,70	-0,27
Quality	1,53	1,15	-0,45	-2,41	-0,17
Appeal	-0,03	-1,21	-1,72	3,15	0,19
Simplicity	-0,73	1,44	0,08	-0,69	0,09
Aesthetics	-1,49	0,55	-2,90	4,35	0,52
Style	0,36	1,52	-0,40	-1,49	-0,01
Harmony	0,14	-0,51	1,32	-1,10	-0,16
Risk	0,14	0,08	1,32	-1,69	-0,16
	-0,97	0,94	1,43	-1,42	-0,02

- distribution depending on judges' levels of experience

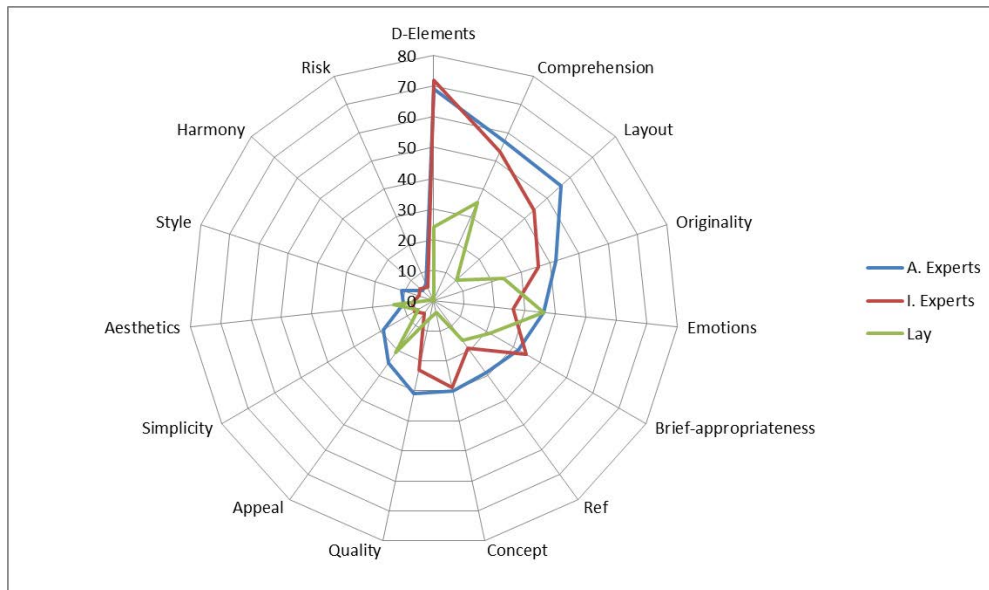


Figure 13: Numbers of occurrences for each ER during the first stage of interviews (visceral choices) within groups of judges with different levels of experience in design field.

Table 37 Standardized Deviates for the number of occurrences of ERs, within the groups of judges with different levels of professional experience.

	A.Experts	I. Experts	Lay	
D-Elements	-0,57	1,62	-1,29	-0,24
Comprehension	0,01	0,06	-0,08	-0,02
Layout	1,00	0,42	-1,89	-0,47
Originality	-0,84	0,08	1,04	0,28
Emotions	-1,14	-1,49	3,44	0,81
Brief-appropriateness	-1,19	1,55	-0,35	0,00
Ref	0,50	-0,87	0,43	0,06
Concept	0,89	1,20	-2,74	-0,64
Quality	1,03	0,79	-2,41	-0,58
Appeal	0,14	-2,50	2,98	0,63
Simplicity	1,33	-0,85	-0,73	-0,25
Aesthetics	-1,44	-1,83	4,27	1,01
Style	1,68	-0,60	-1,51	-0,43
Harmony	-0,16	1,05	-1,12	-0,23
Risk	1,14	0,13	-1,71	-0,44
	2,39	-1,23	-1,68	-0,52

First stage: the assessment based on the visceral reaction

- **distribution depending on judges' backgrounds**

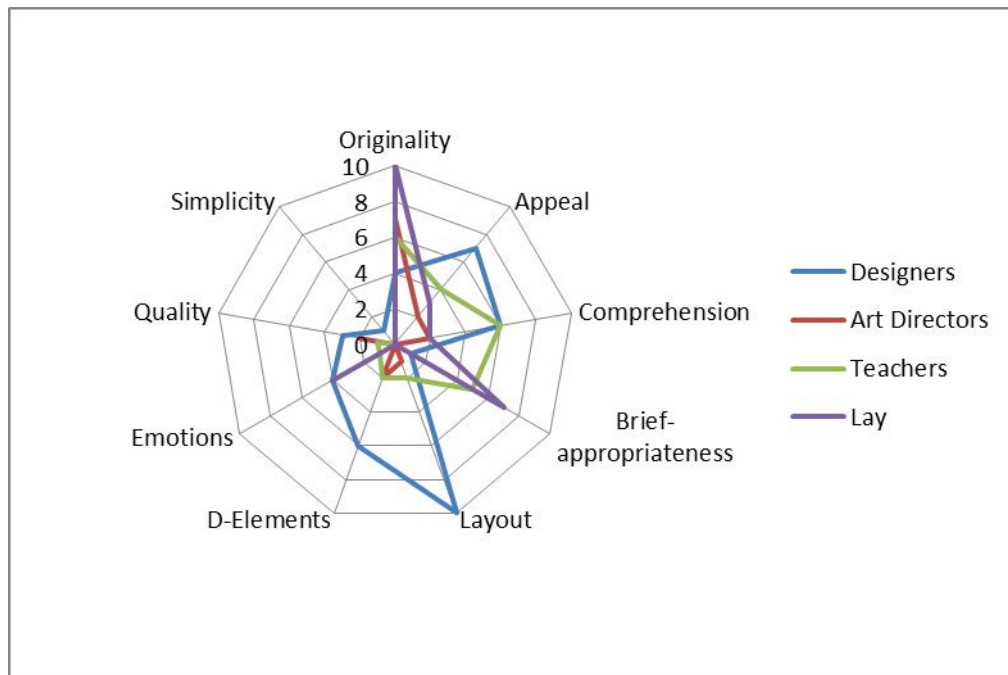


Figure 14: Numbers of occurrences for ERs used during the first stage of interviews (visceral choices), within groups of judges with different backgrounds.

Table 38 Standardized Deviates for the number of occurrences of ERs used during the first stage of interviews, within the groups of judges with different backgrounds..

	Designers	Art Dir.	Teachers	Lay	
Originality	-1,94	1,58	-0,22	1,46	0,87
Appeal	0,38	-0,20	0,05	-0,39	-0,15
Comprehension	-0,02	-0,20	1,07	-0,90	-0,06
Brief-approp.	-1,77	-1,37	1,03	2,27	0,16
Layout	2,29	-0,64	-0,65	-1,75	-0,75
D-Elements	1,14	0,47	-0,28	-1,53	-0,20
Emotions	0,32	-1,14	-0,80	1,30	-0,32
Quality	0,48	1,22	-0,38	-1,19	0,14
Simplicity	1,01	-0,38	-0,49	-0,48	-0,35
	1,90	-0,67	-0,67	-1,20	-0,65

- distribution depending on judges' levels of experience

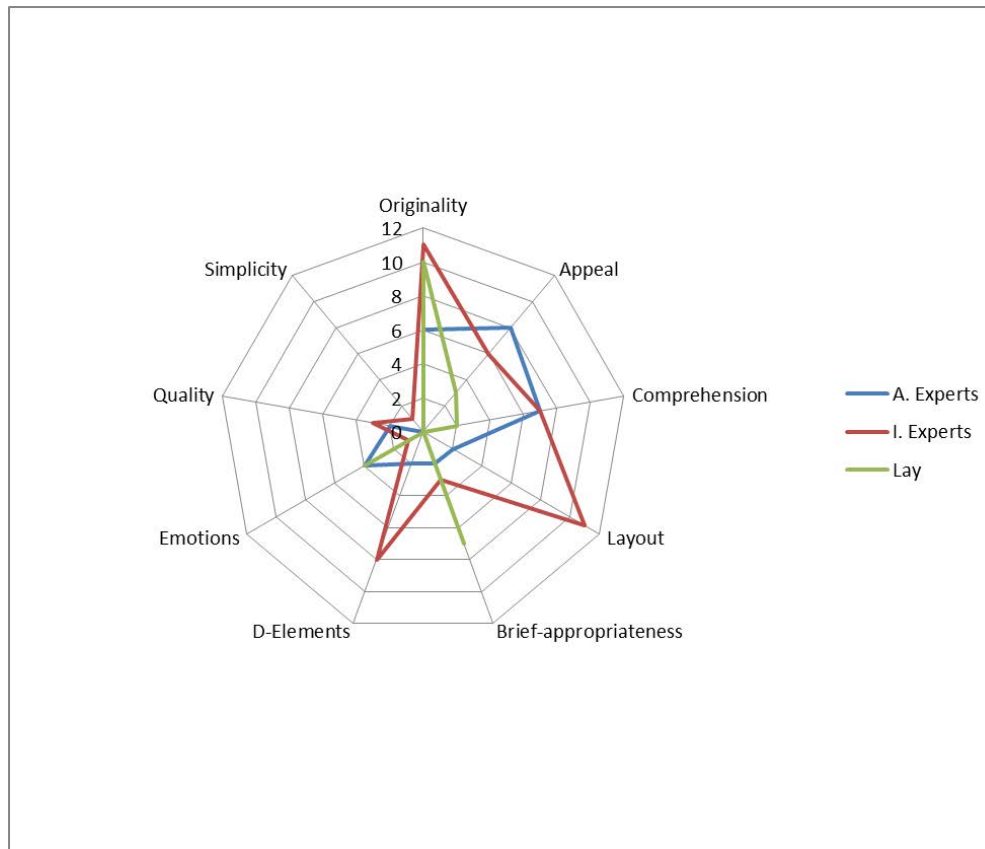


Figure 15: Numbers of occurrences for ERs used during the first stage of interviews (visceral choices), within groups of judges with different levels of experience.

Table 39 Standardized Deviates for the number of occurrences of ERs used during the first stage of interviews, within the groups of judges with different levels of professional experience.

	A. Experts	I. Experts	Lay	
Originality	-0,74	-0,43	1,43	0,27
Appeal	1,28	-0,67	-0,51	0,11
Comprehension	1,00	-0,15	-0,92	-0,07
Layout	-0,96	2,03	-1,75	-0,69
Brief-appropriateness	-0,84	-1,09	2,47	0,54
D-Elements	-0,58	1,56	-1,54	-0,55
Emotions	0,79	-1,55	1,28	0,52
Quality	0,41	0,45	-1,09	-0,23
Simplicity	-0,55	0,79	-0,49	-0,25
	-0,18	0,93	-1,10	-0,35

Second stage: the assessment based on a reflective assessment

- **distribution depending on judges' backgrounds**

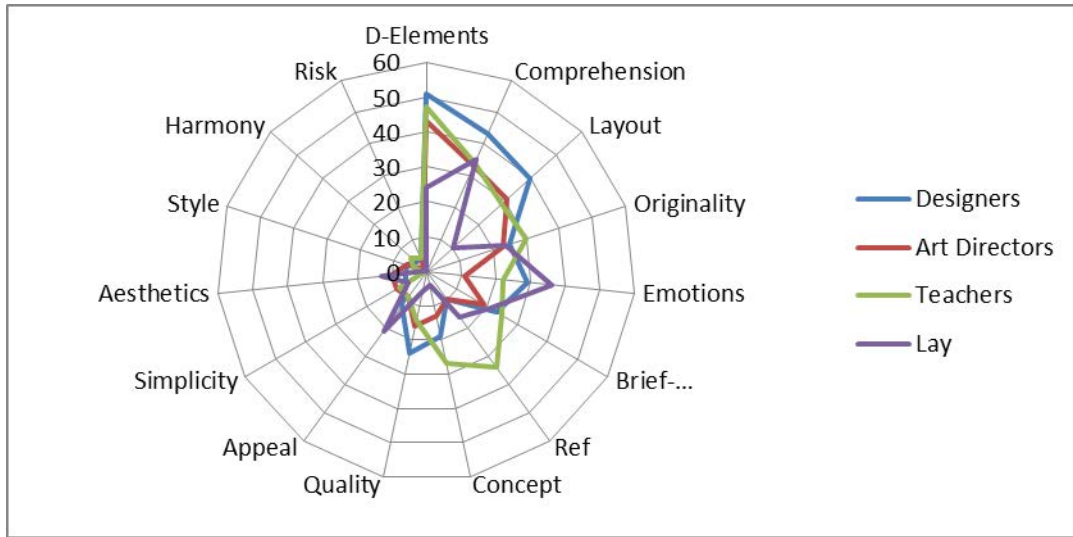


Figure 16: Numbers of occurrences for ERs used during the second stage of interviews (reflective assessments), within groups of judges with different backgrounds.

Table 40 Standardized Deviates for the number of occurrences of ERs used during the second stage of interviews, within the groups of judges with different backgrounds.

	Art					
	Designers	Directors	Teachers	Lay		
Design elements	0,48	0,90	0,22	-1,76	-0,16	
Comprehension	0,17	0,01	-0,95	0,88	0,11	
Layout	1,45	1,20	-0,24	-2,69	-0,28	
Originality	-0,83	-0,03	0,35	0,60	0,10	
Emotions	0,13	-2,39	-0,97	3,46	0,23	
Brief-appropriateness	-0,48	-0,22	0,15	0,63	0,07	
ref	-1,98	-1,19	2,46	0,74	0,03	
Concept	0,18	-0,35	2,31	-2,52	-0,37	
Quality	1,60	0,64	-0,63	-1,83	-0,22	
Appeal	-0,71	-0,76	-1,35	3,19	0,37	
Simplicity	-0,74	1,01	0,06	-0,25	0,08	
Aesthetics	-0,91	1,22	-2,53	2,71	0,49	
Style	0,49	1,09	-0,32	-1,35	-0,09	
Harmony	0,28	-1,05	1,48	-0,95	-0,23	
Risk	0,46	0,32	0,55	-1,51	-0,18	
	-0,41	0,42	0,61	-0,66	-0,04	

- distribution depending on judges' levels of experience

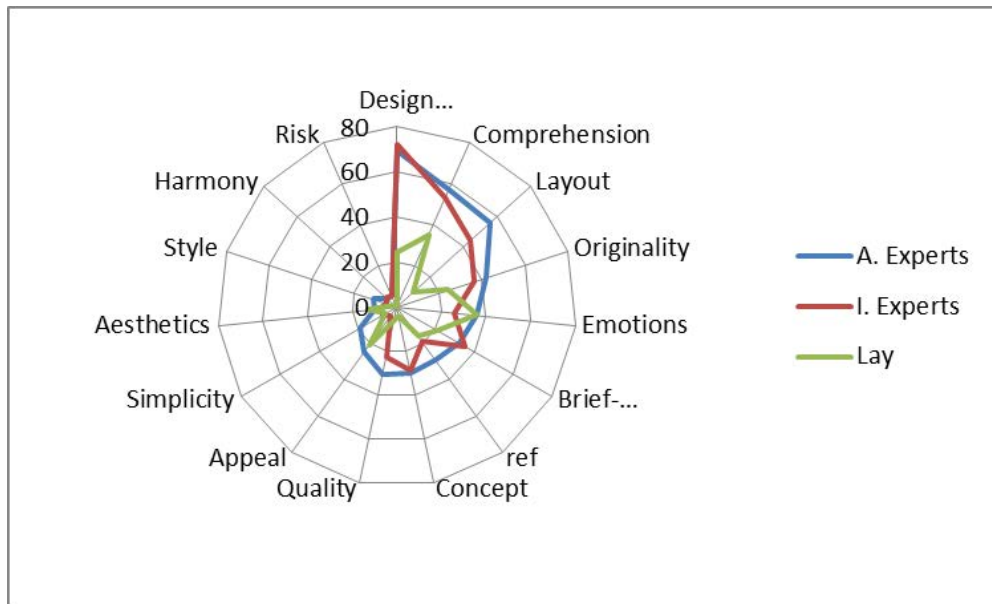


Figure 17: Numbers of occurrences for ERs used during the second stage of interviews (reflective assessments), within groups of judges with different levels of experience.

Table 41 Standardized Deviates for the number of occurrences of ERs used during the second stage of interviews, within the groups of judges with different levels of professional experience.

	A. Experts	I. Experts	Lay	
Design elements	-0,37	1,76	-1,76	-0,37
Comprehension	-0,80	0,21	0,88	0,29
Layout	1,14	0,79	-2,69	-0,76
Originality	-0,39	-0,03	0,60	0,18
Emotions	-1,04	-1,49	3,46	0,93
Brief-appropriateness	-1,04	0,67	0,63	0,26
ref	0,19	-0,78	0,74	0,15
Concept	0,47	1,40	-2,52	-0,64
Quality	0,93	0,37	-1,83	-0,53
Appeal	0,57	-3,08	3,19	0,69
Simplicity	1,34	-1,29	-0,25	-0,21
Aesthetics	-0,86	-1,12	2,71	0,73
Style	1,31	-0,42	-1,35	-0,46
Harmony	-0,11	0,84	-0,95	-0,21
Risk	0,54	0,55	-1,51	-0,41
	1,91	-1,61	-0,66	-0,36

Third stage: choices based on reflective assessment

- **distribution depending on judges' backgrounds**

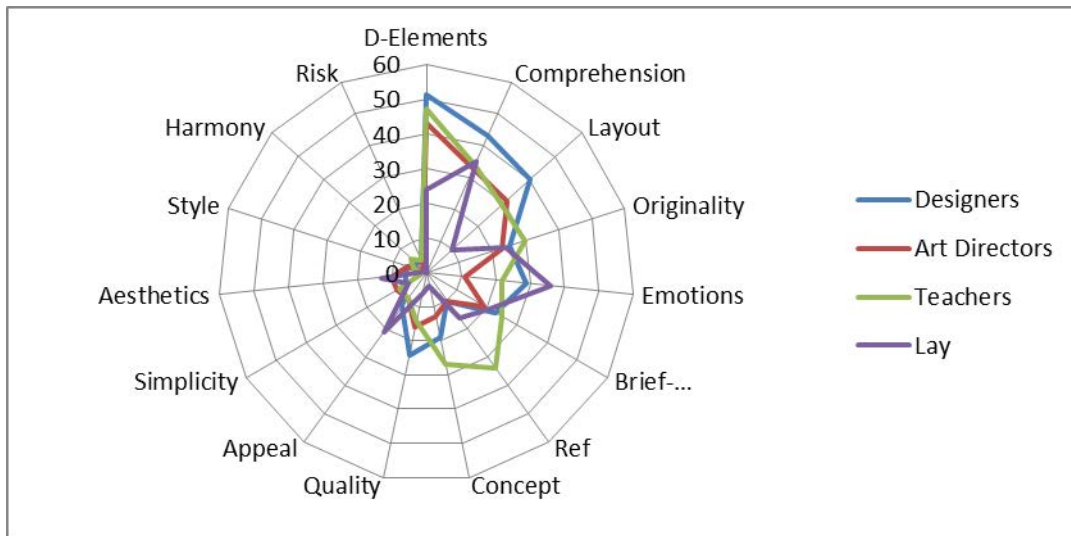


Figure 18: Numbers of occurrences for ERs used during the third stage of interviews (reflective choices), within groups of judges with different backgrounds.

Table 42 Standardized Deviates for the number of occurrences of ERs with positive statements to describe posters chosen during the third stage of the interview (reflective choices), within the groups of judges with different backgrounds.

	Designer	Art directors	Teacher	Lay-people	
Aesthetics	-1,19	0,13	-1,17	2,24	0,002515
Appeal	0,35	-1,06	-0,55	1,42	0,158008
Brief-appr	0,43	0,05	0,14	-0,63	-0,015759
Comprehension	0,05	-0,28	0,14	0,14	0,043246
Concept	-0,92	0,96	1,75	-1,91	-0,131427
D-elements	1,34	0,50	-0,79	-1,16	-0,102242
Emotions	-0,87	-2,11	-0,39	3,71	0,341873
Layout	-0,45	1,93	0,00	-1,77	-0,290988
Originality	0,92	-0,13	-1,23	0,45	0,003406
Quality	1,32	-1,35	1,39	-1,17	0,185063
Ref	-0,36	-0,61	2,24	-1,17	0,101479
Risk	-0,69	1,78	-0,68	-0,68	-0,262609
Simplicity	-0,62	-0,49	0,01	1,17	0,086380
Style	-0,17	1,19	-0,13	-1,07	-0,181558
	-0,86	0,51	0,73	-0,44	-0,062615

- distribution depending on judges' levels of experience

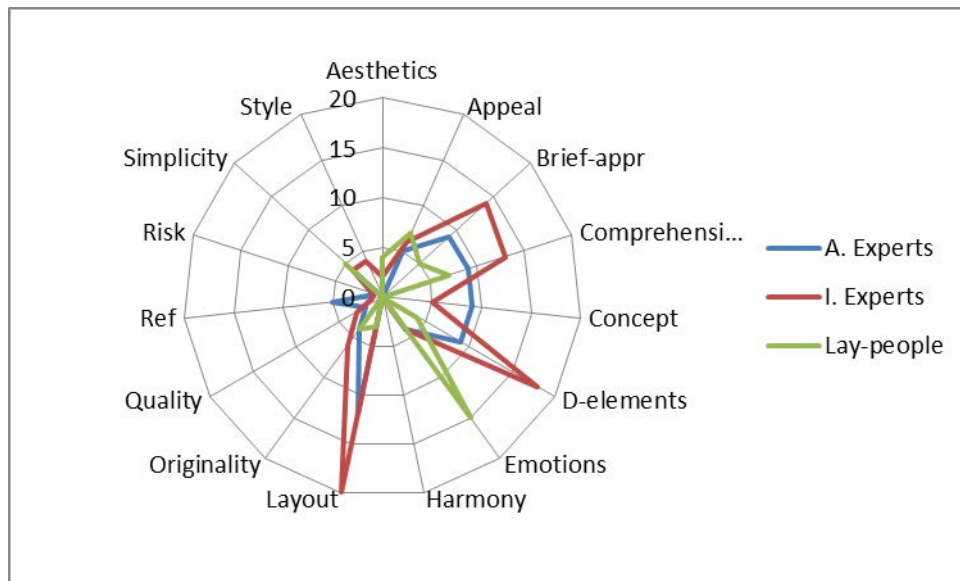


Figure 19: The numbers of occurrences of positive statements used by judges with different levels of professional to describe posters that they chose during the third stage of the interview (reflective choices).

Table 43 Standardized Deviates for the number of occurrences of ERs with positive statements to describe posters chosen during the third stage of the interview (reflective choices), within the groups of judges with different backgrounds.

	A. Experts	I. Experts	Lay-people	
Aesthetics	-1,39	-0,40	2,16	0,37
Appeal	-0,32	-0,70	1,33	0,31
Brief-appr	0,01	0,45	-0,63	-0,17
Comprehension	-0,09	0,04	0,05	0,00
Concept	2,13	-0,48	-1,82	-0,17
D-elements	-0,29	1,15	-1,23	-0,38
Emotions	-1,24	-1,94	4,09	0,91
Layout	0,24	1,14	-1,84	-0,46
Originality	-0,23	-0,08	0,38	0,07
Quality	0,32	0,53	-1,09	-0,25
Ref	2,22	-1,02	-1,19	0,01
Risk	0,45	0,12	-0,69	-0,12
Simplicity	-0,08	-0,73	1,09	0,28
Style	-1,13	1,67	-0,97	-0,43
	0,60	-0,24	-0,36	-0,01

Fourth stage: choices based on personal preferences

- **distribution depending on judges' backgrounds**

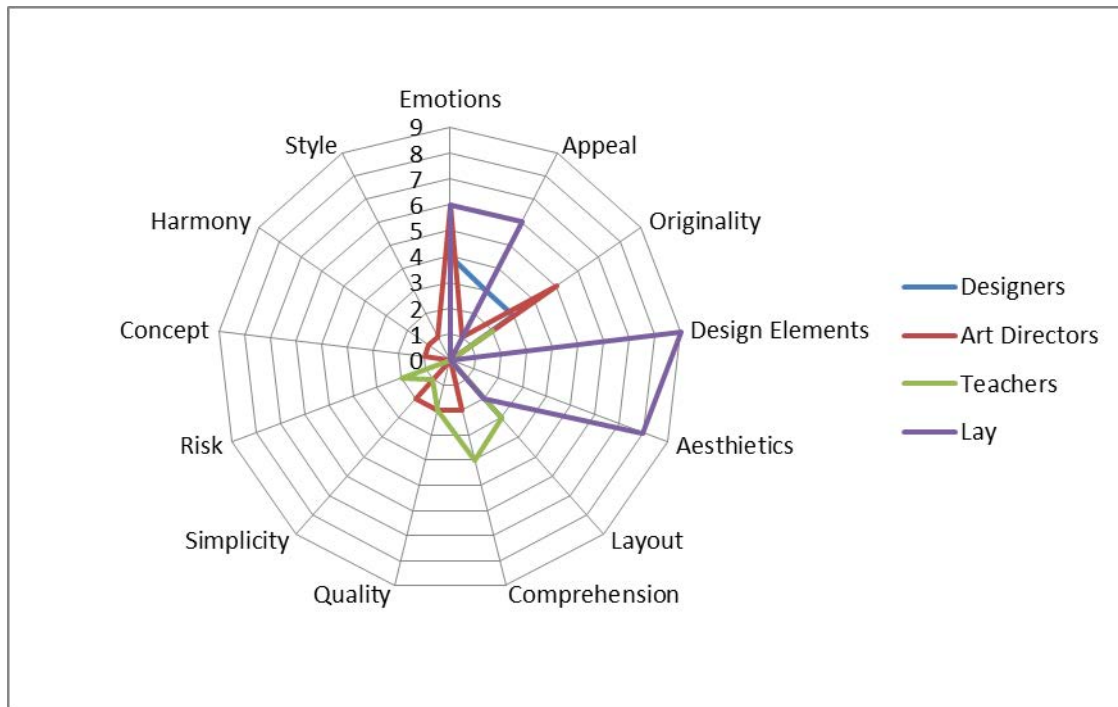


Figure 20: Numbers of occurrences for each ER during the fourth stage of interview (personal preferences) within groups of judges with different backgrounds.

Table 44 Standardized Deviates for the number of occurrences of criteria with positive statements to describe posters chosen during the fourth stage of the interview (personal preferences), within the groups of judges with different backgrounds.

	Art				
	Designers	Directors	Teachers	Lay	
Emotions	0,74	0,73	-1,23	-0,23	0,01823
Appeal	1,08	-1,00	-1,37	1,08	-0,21414
Originality	1,08	1,47	0,09	-1,97	0,66731
D- Elements	-1,21	-1,54	-1,30	2,95	-1,09359
Aesthetics	-1,14	-1,45	-1,22	2,78	-1,03104
Layout	1,49	-1,45	1,22	-0,62	0,64185
Comprehension	-0,99	0,34	2,71	-1,52	0,53701
Quality	-0,81	0,93	1,44	-1,24	0,31927
Simplicity	-0,70	1,37	0,58	-1,08	0,17326
Risk	-0,57	-0,72	2,65	-0,88	0,47862
Concept	-0,40	1,44	-0,43	-0,62	-0,01917
Harmony	-0,40	1,44	-0,43	-0,62	-0,01917
Style	-0,40	1,44	-0,43	-0,62	-0,01917
	-2,23	2,98	2,29	-2,60	0,43926

- distribution depending on judges' backgrounds

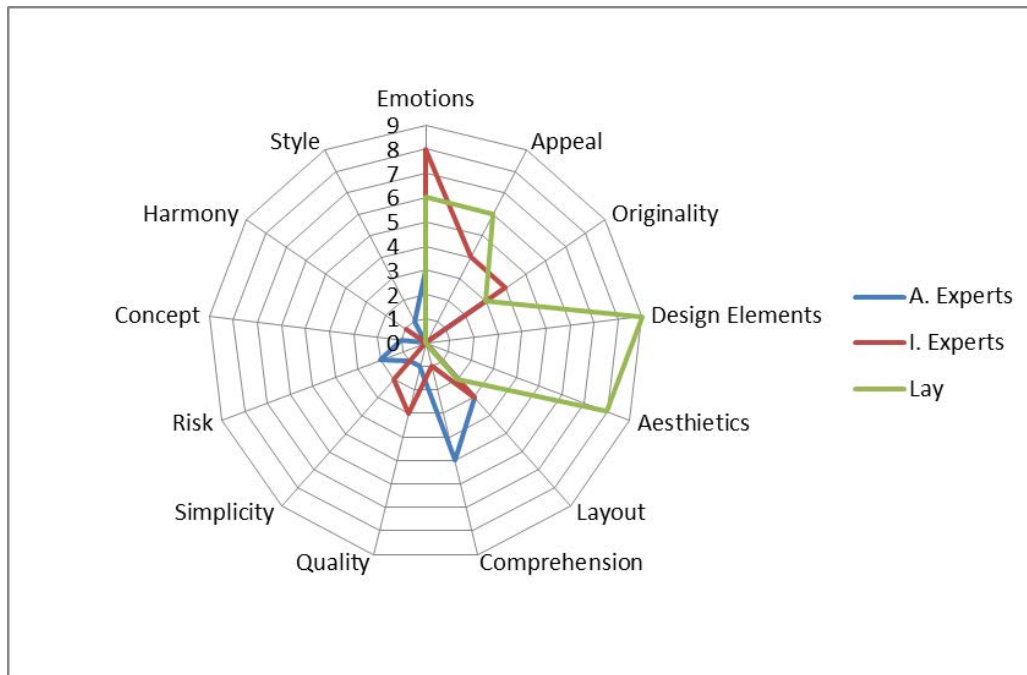


Figure 21: Numbers of occurrences for each ER during the fourth stage of interview (personal preferences) within groups of judges with different levels of experience.

Table 45 Standardized Deviates for the number of occurrences of ERs with positive statements to describe posters chosen during the fourth stage of the interview (personal preferences), within the groups of judges with different levels of professional experience.

	A. Experts	I. Experts	Lay	
Emotions	-0,61	1,05	-0,46	-0,01
Appeal	-1,58	0,42	0,85	-0,32
Originality	0,32	0,42	-0,61	0,13
Design Elements	-1,50	-1,71	2,65	-0,56
Aesthetics	-1,41	-1,61	2,49	-0,53
Layout	0,71	0,25	-0,76	0,20
Comprehension	2,86	-0,68	-1,60	0,58
Quality	0,00	1,49	-1,30	0,19
Simplicity	0,29	1,04	-1,13	0,20
Risk	2,12	-0,81	-0,92	0,39
Concept	1,50	-0,57	-0,65	0,28
Harmony	-0,50	1,18	-0,65	0,03
Style	1,50	-0,57	-0,65	0,28
	3,69	-0,10	-2,74	0,85