The 2nd International Conference on Design Creativity (ICDC2012)
Glasgow, UK, 18th-20th September 2012

DIFFERENCES IN CREATIVE DESIGN ASSESSMENT

A. Wojtczukand[1] and N. Bonnardel[1,2]

[1]Research center in the Psychology of Cognition, Language and Emotion (PsyCLE), Aix-Marseille University, Aix-en-Provence, France
[2]Institut Universitaire de France (IUF) / French University Institute

Abstract: In the context of today's highly competitive society, brands need to attract the audience's attention in ways that are assessed as attractive and creative. It is therefore essential to investigate the process of assessment of creative designs, in order to know which criteria guide judgments on creative productions. In this paper, we especially explore such judgments when they are made by different kinds of actors involved in design. More precisely, the objective of this study is to investigate how judges' creativity assessments are influenced by their professional backgrounds. To analyse the criteria that contribute to judges' understandings of creativity, we asked twenty two judges with different backgrounds (designers, design teachers, art directors and a typical audience) to assess productions in the field of visual communication design. Besides giving an overall creativity score, they were asked to use specific aesthetics, originality, brief-appropriateness and audience-appropriateness. The results allow us to understand how judges' ratings vary depending on their backgrounds. Therefore, this paper contributes to better understanding of criteria that guide assessments of products' creativity.

Keywords: creative design, design assessment, judges' background

1. Introduction

The assessment of creativity has received considerable attention from researchers, not only in terms of psychometric approaches (like those of Guilford, 1967 or Torrence, 1974), but also from a systemic point of view. According to Csikszentmihalyi (1996), creativity cannot emerge from the work of individuals and domain-specific knowledge alone. Creative productions also need to receive social recognition from the experts in the field - the so-called "domain gate-keepers" (Csikszentmihalyi, ibid.). In today's highly competitive society, this subject has become more relevant than ever, especially in the field of visual information design. When different brands offer similar products, customers have to base their choices on visual features, which therefore need both to attract the audience's attention and convey information efficiently. People's product choices are often guided by their initial judgments, based on emotions (Norman, 2004). Hence, visual communication experts often play a crucial role in the success of many companies.

In this context, it is interesting to investigate creativity assessment, taking the different backgrounds of stakeholders or any actors involved in the design process into account. Many authors studying the creative design process have stressed the importance of the evaluation stage, where designers assess each iteration of their project and decide either to press on with it or to look for more interesting ideas (McNeil *et al.*, 1998; Bonnardel, 2000). Others have looked at how teachers assess their students' creative productions in order either to select those of the highest quality or to recommend

improvements (Swann, 2002). In the professional environment, art directors have to select the best version of a given design, anticipating customers' needs and preferences. Marketing studies also focus on the way in which customers assess design, in order to identify those features that attract their attention and prompt them to purchase a specific product (as well as those that do not).

The variability of factors included in creativity assessment mean that it can best be investigated via a multidimensional approach, rather than as a unidirectional construct.

Thus, the main objective of this study is to analyse the way in which judges' different backgrounds affect the criteria that contribute to creativity assessment. More specifically, the analysis of creativity assessment is conducted in the context of visual communication design.

2. Assessing creativity

2.1. Creativity in design

To understand creativity assessment, it is essential to start with the definition of creativity itself (Plucker & Makel, 2010), as it is important to know what people really understand by "creativity" when assessing it.

In the cognitive approach to creativity, many authors agree that a creative product should both be novel (original, unexpected) and appropriate (useful, meet a number of specific constraints (Lubart 1994; Bonnardel, 2006).

Compared with other creative domains, compliance with constraints, or in other words, appropriateness, plays a particularly important role in product design. In visual communication, besides the product's *aesthetic worth*, which is obviously important, the designer also has to take account of the *design brief recommendations* (a document describing the content of the future product, its function, meaning and technical limits) and anticipate the *needs of the target audience* (its characteristics, knowledge and preferences).

At the same time, designers' creativity depends on *originality*, the quality which ensures that their productions will be different from those already known in the domain. As a consequence, there is not one, but many different responses to a given design problem.

2.2. Methods of creativity assessment

According to Baer *et al.* (2004), the ability to assess a product's creativity is one of the best methods of measuring a person's creativity. Indeed, an individual's previous work can be indicative of his or her ability to produce a creative work in the future (Plucker & Mackel, 2010).

Different authors have put forward numerous methods for assessing creative products, including the Creative Product Semantic Scale (Besemer, 1998; Besemer & O'Quin, 1999) and the Student Product Assessment Form (Reis & Renzulli, 1991). These techniques rely on a set of criteria that supposedly evaluate the creativity of a product in an objective way. However, according to some authors, this approach brings with it a "criterion problem", in that we cannot be sure that the chosen criteria are relevant for assessing creativity (Plucker & Makel, 2010).

Teresa Amabile suggested a different method of assessing creativity, called the Consensual Assessment Technique (CAT; see Amabile, 1979, 1982, 1983). This technique consists in selecting appropriate judges - all domain experts - and collecting their independent ratings of creativity. Thus, no specific criteria are given to the judges and the product under assessment is deemed to be creative if there is good independent inter-rater agreement.

Nevertheless, there is considerable debate over how to choose the most appropriate judges. Various authors have studied judge characteristics liable to have an impact on creativity assessment.

White *et al.* (2002), for instance, found that creativity ratings were linked to such variables as age, sex and professional experience in the domain. Hood (1973) showed that these ratings can also depend on the judges' own creative capacity. His results suggested that highly creative judges are more

restrictive in their creative production ratings, giving lower grades than their less creative counterparts.

This can be explained by the level of domain expertise: a novice will have seen fewer examples of creative work than a highly experienced person. Differences in the creativity ratings of judges with different levels of expertise have been studied by many authors, with varying results according to the domain. Thus, creativity ratings by novices and experts are barely correlated in poetry assessments (Kaufman *et al.* 2008), whereas for the assessment of artistic objects (Dollinger *et al.* 2004) or films (Plucker *et al.* 2009), these ratings are highly correlated, even when they are made by judges not just with varying levels of expertise, but also different professional backgrounds.

Another way to consider the differences in judges' approaches to the design, is taking into account the mental models constructed by the different stakeholders, that depend on their backgrounds and previous experiences in the field. For example, following Don Norman's model, designers would develop mental models of products in line with the logic of problem resolution included in design, while users' mental model would be rather based on previous user-experiences (Norman, 1988). It is also interesting to understand the differences between the mental models of the experts or "domain gate-keepers" (Csikszentmihalyi, 1996) and those of non-experts. Even if their judgment of creativity can be highly correlated, their mental models of what is important in a creative product could vary.

The present study sought to understand how judges' creativity assessments are influenced by their professional backgrounds. More specifically, we studied the criteria that contribute to judges' possibly different understandings of creativity in the field of design.

2.3. Creativity assessment criteria

Research has been conducted with a view to identifying the most relevant assessment criteria in different fields of creative design. According to Sarkar and Chakrabarti (2008), who brought together existing definitions of creativity, to assess the creativeness of newly designed products, the judges must be able to assess the originality (or novelty) and usefulness of these products.

Many authors studying creativity used different criteria of creativity assessment and most of them included "originality" and "usefulness", more often called "appropriateness", into their criteria sets.

The **originality** is described as innovation (Kreitler & Casakin, 2009), the product being unusual and unexpected compared to other ideas (Shah et al, 2003; Bonnardel, 2006). Sometimes authors consider separated fields of originality, for example Shah et al, (2003) distinguish novelty concerning function of product and its physical principles.

The **appropriateness** criterion can have several understandings. Some authors take into account the degree to which the product fulfills different specifications, they employ terms such as "appropriateness" (Sarkar & Chakrabarti, 2008; Caroff, 2008), "quality" (Shah et al, 2003), or relevance (Cropley & Cropley, 2005). Some authors focus on the degree to which the product can be **attractive for the user**, for functional products' assessment, they employ terms like "usefulness" (Chakrabarti & Khadilkar, 2003; Kreitler & Casakin, 2009) or "ergonomics" (Cross & Dorst, 2001). In some approaches, the affect transmitted by a design is also taken into consideration (Horn & Salvendy, 2009).

This is in line with marketing approaches: "emotional design" (Norman, 2004) or "Kansei engineering" (Yan *et al.* 2008), which emphasise the need to take the emotional and aesthetic dimensions into account. **Aesthetics** criteria appear in numerous creativity assessments studies, under terms such as "elegance" (Cropley & Cropley, 2005), "mastery of skills concerning the aesthetics of the design representation" (Kreitler, Casakin, 2009) or as very specific criteria related to shapes, colours, rhythm, order, etc. (Amabile, 1982; Demirkam & Afacan, 2012).

Over previous studies investigated relationships among several criteria, Cross and Dorst (2001) found that the "ergonomics" judgment of the industrial designs correlated most heavily with the overall judgment, opposite to "creativity" (understood as originality) which correlated the least. Some criteria can affect each-other, like aesthetics can affect the apparent usability of an interface (Tracinsky *et al.* 2000). Caroff (2008) studying the assessments of advertisements found an interaction: overall

creativity ratings were more favourable when the level of originality was high, but when a certain appropriateness threshold was exceeded, productions were considered as uncreative independently of their originality level.

For their part, Wojtczuk and Bonnardel (2011) explored how different criteria (originality, aesthetics, functionality and marketability) influence the assessment of innovative designs for everyday objects (in this case, computer mice). More specifically, this research focused on differences between different groups of judges (designers, design teachers, retailers and users). Results highlighted a number of similarities between users and retailers, and between designers and design teachers. This is in line with Norman's statement about the difference between designers' and users' mental models of systems (Norman, 1988), but it also tells us something about the mental models themselves. The users and retailers based their assessments on marketability and functionality, whereas the designers' and design teachers' assessments were based mostly on aesthetics.

3. Experiment

Following on from the research by Wojtczuk and Bonnardel (2011), the present experiment was designed to investigate differences in the creativity assessments made by judges from different backgrounds, in the context of visual communication design (graphic posters). We believe that this type of design is more information-centred and that its aesthetic appeal is important to the overall assessment.

Our aim was to ascertain how the implicit understanding of creativity of the different judges, all potentially involved in the process of poster design, diverged according to their different backgrounds: designers involved in creation; art directors involved in approval; design teachers specializing in design training; and an audience (potential target). We asked these different judges to assess a series of posters, rating them on overall creativity plus several specific criteria relevant to visual communication design. To find out which criteria contributed most to the overall creativity assessment, we investigated the correlations between the creativity scores and the scores for these other criteria.

Our general hypothesis was that background has an impact on the way that judges assess a visual communication design. We expected to find differences between the criteria that guided the creativity assessments made by judges from different backgrounds.

Our more specific hypotheses were that:

- Following the Norman's differentiation between users' and designers' mental models (Norman, 1988), we expected to find differences of judgments between these two groups. In line with Wojtczuk and Bonnardel (2011) findings, we expect that art directors share the audience criteria of judgment, while design teachers share those of designers.
- Still following Normans' model, we expected the designers' assessments to be guided more by originality (wondering whether they were capable of having a similar idea), while we expected the audience to be more interested in the posters' aesthetics and audience appropriateness.
- Furthermore, we expected that art directors and teachers would focus more on the posters' appropriateness, as we perceived them as fulfilling the role of "domain gate-keepers", supervising the productions' match with current trends (Csikszentmihalyi,1996).
- As we believe that visual appeal is very important to poster design, we predicted that aesthetics would be an important criterion for all the judges.

4. Methods

4.1. Participants

We selected 22 men and women from different backgrounds as our judges. We wanted to compare the judgments of professional designers, art directors, design teachers and individuals who could represent the posters' potential audience.

Thus four groups of participants were constructed as follows:

- five professional designers, with 9.4 years of experience on average;
- five design teachers, with 12.6 years of teaching experience on average;
- five art directors (who make decisions about style and overall visual appearance, and give directives to designers), with 10.8 years of professional experience as art directors on average;
- seven participants with no experience in the design domain, who represented a potential audience for the posters: four men and three women, aged 18-65 years (balanced age and sex mix).

4.2. Procedures

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 theme, the goal being to inform the potential audience about a specific event.

In this study, we needed to adapt the criteria to the field of visual communication design. With the help of a professional designer, we decided on the four most appropriate criteria for measuring the creativity of the posters in this field. We wanted to include the two criteria that correspond to the definition of creativity (i.e., originality and appropriateness), but as "appropriateness" seemed insufficiently clear, we decided to make it more specific. We therefore decided to ask the judges about the posters' brief-appropriateness, and gave them access to the design brief that had been given to the posters' creators (the design brief described the content, the problematic of the subject and the format of presentation). In addition, we added a criterion of audience-appropriateness, which corresponds to supposed attractiveness of the product for the targeted audience. Moreover, aesthetics also appeared to be important in the context of visual communication design assessment.

The assessments were made on line by accessing a specific experimental website. To access this website, each judge was given an individual ID and password. 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 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 finishing, the composition, the aesthetical pleasure felt while looking at the product;
- originality corresponds to the innovative side of the product;
- audience 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.

5. Results

To analyse the scores awarded by the four groups of judges, we conducted an analysis of covariance (ANCOVA). When we controlled for each of the four criteria (originality, aesthetics, brief-appropriateness and audience-appropriateness), we failed to find any significant effect of judges' background on creativity judgments, $F(1, 454) = 1{,}54$, $p = .2$.

Nevertheless, linear regression analyses by group revealed that the four criteria had differing effects on the creativity assessments of each group of judges.

- For designers, we found highly significant effects of originality, b = .35, SD = .08, β = .37, p < .001, and audience-appropriateness, b = .33, SD = .08, β = .36, p < .001.
- For design teachers, we found highly significant effects of brief-appropriateness, b = .4, SD = .08, β = .37, p < .001, and originality, b = .26, SD = .06, β = .26, p < .001.
- For art directors, we found a highly significant effect of brief-appropriateness, b = .52, SD = .07, β = .55, p < .001.
- For the audience, we again found a highly significant effect of brief-appropriateness, b = .28, SD = .09, β = .11, p < .001.

When we examined the correlations between the different criteria, we found that all the criteria were positively correlated.

To provide a clearer picture of the distribution of mean scores assigned by the different groups of judges for the different criteria, they are illustrated in the figure below (see Figure 1.).

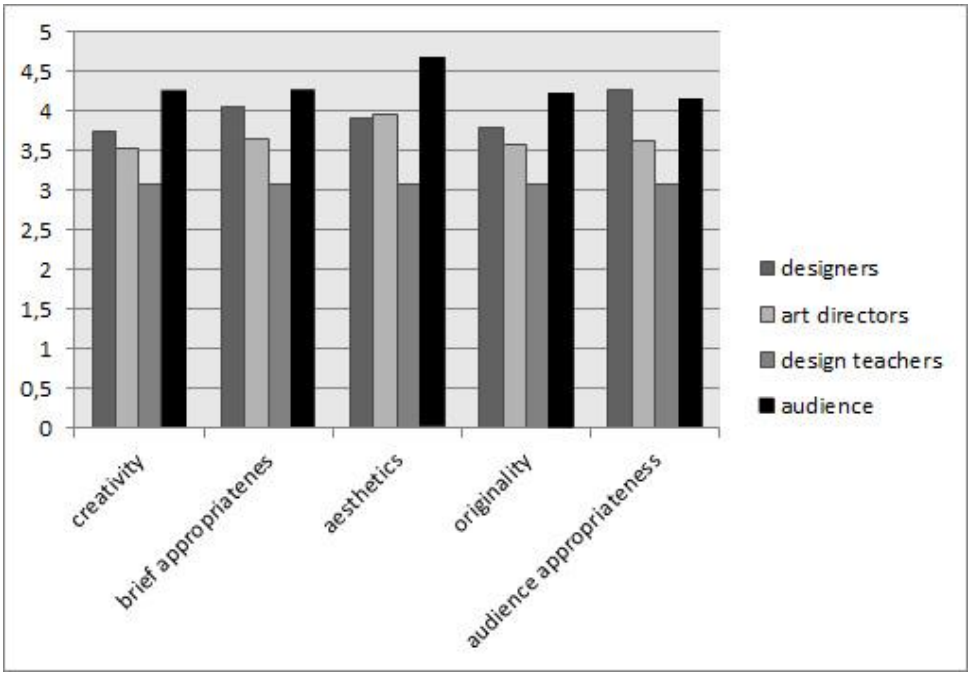

Figure 1. Distribution of mean scores assigned by the four groups of judges according to different criteria.

The figure 1 shows that the highest scores were attributed by the audience, especially for aesthetics, while the lowest scores were assigned by the design teachers, chiefly for originality. The designers tended to award the highest scores for audience-appropriateness.

6. Discussion

The global results failed to reveal any significant difference between the creativity scores awarded by the judges. This is in line with Amabile's consensual assessment method, which is based on the assumption that judges give similar scores for creativity, with no particular differences between groups. Nevertheless, when we analysed the judges' assessments in greater detail, we found that these overall creativity assessments were associated with different criteria.

We found that brief-appropriateness was the criterion that guided most of the judges in their creativity assessments (audience, design teachers and art directors). This partially contradicted our hypothesis, as we had expected that only the art directors and teachers would focus on this criterion. Brief-appropriateness therefore seems to be of more general importance, even for the nonspecialists (audience), who are not normally interested in design briefs. This could be due to the central function of content in visual communication design, which was carefully set out in the brief. Content should

inform and convey a specific message. Thus, in our case, the assessment consisted partially in validating the relevance of the posters' content to the message described in the brief.

We found that the originality criterion contributed to the designers' and design teachers' assessments of creativity, but not to those of the two remaining groups (audience and art directors). This confirmed our hypothesis of a distinction between judges with a "creative" profile and those with a "user" profile, as in Norman's model (Norman, 1988). Thus, the designers and teachers seemed to focus on the creative quest, while the audience and art directors seemed to focus on ease of comprehension by the targeted audience.

The designers appeared to form a separate group from the others, as theirs was the only group for which brief-appropriateness was not the most important criterion. For them, originality seemed to be the criterion that contributed the most to their creativity assessment (in line with our hypothesis), along with audience-appropriateness. For this group, it may be important to go beyond the design brief's recommendations, seeking out original ideas and thinking more about audience needs than about the basic functions of the designed object.

Our hypothesis was based on the assumption that creativity assessments of visual communication designs rely heavily on aesthetic values for all judges, but more especially for the audience. Results, showed, however, that aesthetics was not the most important criterion for any of the groups. It may have been so obvious that the posters needed to be attractive that the judges did not treat it as a criterion for assessment, but instead as something in addition to the whole creative concept of the poster. As for our prediction that the audience would display a preference for highly aesthetical productions, we found that even though the aesthetics criterion did not strongly influence their overall creativity assessment, they nonetheless awarded their highest scores for aesthetics. This may be connected with the general tendency for nonexperts to rate productions from a field in which they have little knowledge more highly, which would confirm the findings of Plucker *et al.* (2009) and Dollinger *et al.* (2004), concerning the influence of expertise on the creativity rating.

We also observed that in contrast to the audience's high ratings, the design teachers had a tendency to attribute far lower scores. This confirms the results reported by Hood (1973), who concluded that the greater a judges' creative capacity, the more restrictive they will be in rating creative productions.

In conclusion, the present study highlighted the usefulness of exploring individual differences between judges in the process of assessing creativity, especially concerning their professional backgrounds. It had several limitations, mainly due to the small number of judges. Although we obtained a sufficient amount of data, by considering every single judgment as an independent set of scores, it would be worthwhile confirming our findings with a larger sample of participants. Nonetheless, our research emphasises the fact that the processes by which judges assess products in the field of visual communication design differ depending on their background, even if they reach similar overall creativity scores. Further studies are needed to explore this assessment process in greater detail and analyse the judges' verbalisations in order to understand the reasons behind their scoring better.

Acknowledgement

A part of this study was done within the CREAPRO research contract funded by ANR (Contract ANR-08-CREA-038). We would like to thank Marylaine Pupet and her students from the AxeSud school, as well as Thomas Arciszewski, for their contributions to this work.

References

Amabile, T. M. (1979). Effects of external evaluation on artistic creativity. *Journal of Personality and Social Psychology,* 37, 221–233.

Amabile, T. M. (1982). Social psychology of creativity: A consensual assessment technique. *Journal of Personality and Social Psychology, 43,* 997–1013.

Amabile, T. M. (1983). *The social psychology of creativity*. New York: Springer-Verlag.

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.

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.

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 [Creativity and design. Cognitive and ergonomic approach]*. Marseille : Solal Editions.

Caroff, X. & Besançon, M. (2008). Variability of creativity judgments. *Learning and Individual Differences, 18(4),* 367-371.

Chakrabarti, A., & Khadilkar, P.(2003). A measure for assessing product novelty, *Proceedings of the International conference on Engineering Design, 2003, Stockholm, 2003.*

Csikszentmihalyi, M. (1996). *Creativity: Flow and the Psychology of Discovery and Invention.* New York: Harper Perennial.

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.

Dorst, K., & Cross, N. (2001). Creativity in the design process: co-evolution of problem and solution. *Design Studies, 22*(5), 425-437.

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.

Guilford, J. P. (1967b). *The nature of human intelligence.* New York: McGraw-Hill.

Hood, R. W. (1973). Rater originality and the interpersonal assessment of levels of originality. *Sociometry, 36,* 80-88.

Horn, D., & Salvendy, G. (2009). Measuring consumer perception of product creativity: impact on satisfaction and purchasability. *Human Factors and Ergonomics in Manufacturing, 19*(3), 223-240.

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.

Kreitler, S., & Casakin, H., (2009).Self-perceived creativity: the perspective of design. *European Journal of Psychological Assessment, 25*(3), 194-203.

Lubart, T. (1994). Creativity. In R.J. Sternberg (Ed.), *Thinking and Problem Solving* (pp. 289-332). New York: Academic Press.

McNeill, T., Gero, J. S., & Warren, J. (1998). Understanding conceptual electronic design using protocol analysis. *Research in Engineering Design, 10,* 129-14

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.

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.

Plucker, J. A., Makel, M. C. (2010). Assessment of creativity. In J. C. Kaufman & R. J. Sternberg (Eds.), (pp. 48-73). *The Cambridge Handbook of Creativity.* Cambridge: Cambridge University Press.

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.

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.

Shah, J., Vargas-Hernandez, N., Smith, S. (2003). Metrics for measuring ideation effectiveness. *Design Studies, 24,* 111-143.

Swann, C. (2002). Nellie is Dead. *Art, Design & Communication in Higher Education, 1(1),* 50-58.

Torrance, E. P. (1974). *Torrance Tests of Creative Thinking: Norms-technical manual.* Bensenville, IL: Scholastic Testing Service.

Tractinsky, N., Shoval-Katz, A., & Ikar, D. (2000). What is Beautiful is Usable. *Interacting with Computers, 13,* 127–145.

White, A., Shen, F., & Smith, B. L. (2002). Judging advertising creativity using the creative product semantic scale. *Journal of Creative Behavior, 36,* 241−253.

Wojtczuk, A., & Bonnardel, N. (2011). Designing and assessing everyday objects: Impact of externalisation tools and judges' backgrounds. *Interacting with Computers, 23(4),* 337-345.

Yan, H., Huynh, V., Murai, T., & Nakamori, Y. (2008). Kansei evaluation based on prioritized multi-attribute fuzzy target-oriented decision analysis. *Informatio Sciences, 178,* 4080–4093.