

BEYOND CREATIVITY ASSESSMENT: COMPARING METHODS AND IDENTIFYING CONSEQUENCES OF RECOGNIZED CREATIVITY

Dagny Valgeirsdottir¹ and Balder Onarheim²

¹Copenhagen Institute of NeuroCreativity, Copenhagen, Denmark

²Department of Management Engineering, Technical University of Denmark, Kgs. Lyngby, Denmark

Abstract: Can people recognize and appreciate design creativity in products? It has previously been shown that creativity influences willingness to purchase products. Those results served as the inspiration for this study, however, it was of interest to investigate whether using a different research approach would yield similar results. Thus the Consensual Assessment Technique (CAT) (Amabile, 1982) was adopted. Participants were asked to assess creativity level, technical advancement and aesthetic appeal, as required when applying CAT, adding purchasability to investigate appreciation of creativity, which is outside the usual CAT frame.

Despite the expansion of CAT a high interrater agreement existed for each attribute indicating that CAT was reliable. This study could, however, not reproduce the previous findings of a relationship between creativity and purchasability of design products. Aesthetic appeal was the only attribute shown to predict purchasability, a relevant finding for designers and managers alike.

Keywords: *Design creativity, creativity assessment, consensual assessment technique, CAT, creativity, aesthetic appeal, purchasability, product assessment.*

1. Introduction

Creativity is an intricate concept, which for many is raveled with mystery. Historically, it has been suggested that although people have difficulties pointing out exactly what product creativity is, it seems to be recognizable by most (Amabile, 1982). Several researchers have studied the subject of defining creativity, e.g. Amabile and Runco and Jaeger, but still there exists little knowledge about what it is exactly that people respond to as being creative, and further what the consequences of recognizing it are. In this study the focus is on the consequences of creativity, i.e. if people can recognize creativity in design objects will they consequently appreciate it so they could imagine acquiring the product?

The focus on appreciation of creativity lead to the interest of further investigating the relationship between product creativity and willingness to purchase, previously identified by Horn and Salvendy (2006b; 2009). They conducted studies to investigate product creativity, subsequently developing and applying the Product Creativity Measurement Instrument (PCMI). For the purposes of this study it was of interest to apply a different creativity assessment method than PCMI, namely the

Consensual Assessment Technique (CAT) (Amabile, 1982) to assess if a similar relationship between design creativity in products and purchasability could be identified. To the best of our knowledge CAT has not previously been used with the purpose of investigating that relationship, although it is still a widely accepted and applied creativity assessment method (Lau et al., 2013; Lebuda and Karwowski, 2013; Stanko-Kaczmarek, 2012). The benefit of establishing the applicability of CAT in the context of appreciation of creativity is that it is an instrument that is simple to deploy because of its subjective nature.

The fundamental difference between the two assessment methods is that PCMI utilizes rating scales with semantic pairs to assess predetermined criteria of creativity, thus classified as a rating scale measurement tool (Horn and Salvendy, 2006b). Participants working within the framework of PCMI are asked to assess objective criteria that are assumed to describe what creativity constitutes of. CAT is a simpler, yet very reliable (Hennessey, Amabile and Mueller, 2011), instrument and differs from PCMI as it assumes that judgments in regards to creativity can only be subjective, thus not including predetermined criteria, but rather relying on the subjective assessments of participants working within its framework. Further, CAT is applied to assess tasks that are open-ended, which allows for more flexibility and novelty in responses, utilizing heuristic activities (Hennessey, 1994). There are five requirements to be met when utilizing CAT, which were adopted for the purposes of this study. However, the standard procedure was expanded and modifications were made to the typical setup.

2. Theoretical Background

In recent years the understanding of creativity has been receiving greater emphasis in an organizational setting as creativity is increasingly being regarded as a factor that can lead to economic success (Sawyer, 2012). However, although creativity is a valuable entity in terms of innovation and product design, there has been a confusion surrounding the concept, historically deterring organizations and individuals from realizing its importance (Sternberg, 1999). Throughout the history of creativity research, there have been made numerous attempts to form a definition of it. A proper definition of a concept is the foundation for successful assessment of it, as it serves as the basic construct and in an effort to gain an overview Mayer (1999) made an effort to gather, from the abundance of definitions put forth, the most common denominators historically used to define creativity. The conclusion was that “original and useful” were the most common denominators, though often put forth in the forms of synonyms such as “novel and appropriate”, “new and valuable” etc. It has also been suggested that in addition to the two aforementioned criteria, a product or response can only be considered creative to the extent that it is publicly recognized as such (Amabile, 1982). However, as Runco and Jaeger (2012) concluded, Stein (1953) was first to conceptualize what is considered to be the standard definition of creativity, choosing to define creativity as something novel and useful.

Historically it has been of great interest to assess creativity and many assessment methods have been constructed and implemented. Creativity assessment has moreover been studied within the field of computational creativity and researchers (e.g. Brown, 2008) have proposed that computer science can contribute to it due to increased precision through the use of computable constraints. In this article the focus will, however, be on the two previously introduced methods, CAT and PCMI, both defined as product creativity assessment methods. In the work of Horn and Salvendy terms such as “willingness to purchase”, “willingness to pay” and “purchasability” have been used. The last term was adopted here as it was considered to be a product attribute entailing appreciation, in terms of wanting to acquire, rather than specifying purchase intent, as that concept entails a prediction of behavior. Research has historically suggested that it is skeptical to predict behavior, thus no such propositions will be made in this study.

The assessment methods originate from two different standpoints of product creativity assessment, however both are based on similar operational definitions of creativity but have different approaches to assessing it, which will now be specified. Horn and Salvendy (2006a; 2006b; 2009) proposed questions concerning how possible buyers perceive products as creative, how much a product value is enhanced by that perceived creativity and how it influences willingness to buy. Those questions were the catalyst for the development and testing of PCMI (Horn and Salvendy, 2009). The

model is originally adapted, amongst other, from the creative product analysis matrix (CPAM) (Besemer and Treffinger, 1981), and assumes that creativity can be assessed using universal criteria based on the three dimensions of affect, novelty and importance. Two additional assumptions are made which are that product creativity assessment is subjective as it relies on the judge and the context it exists in, as well as that product creativity adds value to the overall assessment of a product (Horn and Salvendy, 2006b). PCMI was shown to predict consumer attitudes and the affect dimension of creativity was shown to account for much of the explained variance of purchase intentions and customer satisfaction, indicating the importance of creativity when considering willingness to pay (Horn and Salvendy, 2009). The significance of those findings were that it was shown that the affect dimension had equal effect on creativity assessments as the novelty dimension, suggesting that both dimensions should be considered when evaluating creativity. However, in terms of possible buyers of the products the affect dimension was shown to have greater effect on willingness to pay and satisfaction, which entails important indications for organizations in terms of an enhanced competitive advantage (Horn and Salvendy, 2009). A separate study (Christensen, Kristensen and Reber, 2009) further supported that creativity significantly predicted willingness to pay. However, this study was not conducted under a specific creative assessment framework.

CAT (Amabile, 1982) is an interesting creativity assessment tool for a number of reasons, one of them relating to a fundamental debate in creativity research and theories. The debate is regarding the nature of creativity and the issue of domain generality versus domain specificity. CAT has been said to be the only method not tied to any specific theory of creativity, thus neutral in regards to the ongoing debate about domain specificity (Baer and McKool, 2009). Furthermore, it is a subjective assessment tool, which has been used for over three decades to investigate the creativity of products and other creations. Researchers that utilize the CAT assume that creative assessments can only be subjective and that participants assessing products within the framework of CAT will, despite not being provided with objective criteria to define creativity, recognize it and can largely agree in their ratings (Hennessey et al., 2011). Thus it is assumed that products or creations can only be creative if the appropriate observers or judges find them creative. Those deemed as appropriate judges are persons that have knowledge within the domain in which the products or creations originate from (Amabile, 1982; Hennessey, 1994). It is on average performed with ten judges, but ranging from two to 40 judges. The judges observe the product or creation to be assessed and rate it, usually on a scale of zero to five. However, any scale could be used, though preferably not with less than three options (Baer and McKool, 2009). The quality of the method; that it relies on subjective assessments, eliminates the difficulty of producing definite criteria for creativity, and allows the assessment to be more alike everyday assessment of creativity within specific domains (Hennessey 1994). It is claimed that the lack of criteria makes the instrument more robust, as it does not rely on researchers choosing the right criteria. A wealth of data exists about the reliability of the approach, however there is not as much to be found on what it is exactly that the judges are responding to (Hennessey, 1994). Research using CAT has proven that although creativity has historically been hard to define, it is certainly something that people recognize. In addition to that, people who have knowledge within a specific domain seem to be able to agree upon what is creative in that domain (Amabile, 1982).

There are five requirements to be met in order to utilize CAT (Amabile, 1982). First, all judges that take part in the assessment procedure must be familiar with the domain in question. Secondly, it is a requirement that all judges make their assessments independently, without guidance and criteria, and that no communication takes place between the judges when they are contemplating on their ratings of creativity. The third requirement is that the ratings must be done relative to the other products in the procedure, instead of rating them against some ultimate standard, which could lead to the result of some being rated as non-creative if rated relative to some highly creative work. Fourth is the requirement of judges rating the products in random order to prevent a bias of the methodological order of the products. The final requirement is that if CAT is being applied to a product for the first time, judges should be asked to assess other features, in addition to creativity level, e.g. technical advancement and aesthetic appeal, in order to assess how these other features relate to, or are independent of, the creative judgment (Amabile, 1982).

In a recent study Kaufman and Baer (2012) came to a conclusion that the use of experts is what gives CAT its validity, in addition to reliability acquired through interrater agreement. Lu and Luh (2012) compared PCMI and CAT in a separate study, in terms of raters and applicability, and came to the conclusion that non-expert raters had higher interrater reliability and explanatory power than experts within both frameworks. Furthermore it was suggested that PCMI was applicable to similar tasks as CAT, its items yielding higher explanatory power for creativity scores using CAT (Lu and Luh 2012). The use of experts when deploying CAT will be discussed further in the upcoming method chapter.

2.1. Hypothesis

A hypothesis was formed based on the theoretical background. H1 is drawn from the theoretical background concerning Horn and Salvendy's (2006b; 2009) studies where a relationship between creativity and purchasability was identified. The results concerning this hypothesis will show whether a similar relationship can be identified applying CAT.

H1: Design creativity in products can be shown to predict purchasability when applying the CAT.

The hypothesis was attempted, supported by rejecting the hypothesis:

H0: Design creativity in products cannot be shown to predict purchasability when applying the CAT.

Another catalyst for the hypothesis were results from a previous study (Valgeirsdóttir, Onarheim & Gabrielsen, 2014), which indicated a relationship between creativity and purchasability using a research design similar, yet not identical, to the CAT.

3. Method

For the purposes of testing the hypotheses an experiment was designed conforming to the framework of CAT. It was conducted in an online format, though in a classroom and all participants taking part at the same time, without being able to confer. Participant demographics were master students of mostly Scandinavian descendant within the age group of 23-28. The students were a mix of engineering students, mainly engineering designers. The task of the 22 participants was to randomly assess 15 smartphone accessories, as that product category was assumed to be knowledgeable to the participant group. The experimental setup was inspired by Hennessey's (1994) CAT setup, i.e. participants were first asked to review photos of all the products with a short description text to produce a baseline for the assessment. Consequently the product attributes to be assessed each appeared on a separate slide, first creativity level, next technical advancement, then aesthetic appeal and lastly purchasability. The products were randomized per participant.

3.1. Modification to CAT setup

All requirements of CAT were met in the experimental setup; however, one modification was made to the typical setup, i.e. participants were asked to assess purchasability of the products. This modification was implemented in order to assess whether willingness to acquire could be identified using CAT.

3.2. Participants

The first requirement of CAT is to ask appropriate judges to assess products or responses. Those deemed appropriate as judges, or participants, are those familiar with the domain from which the product or response originates. It is not required to be familiar with the process of producing the item to be assessed, but to be familiar with the creative product, or result of that process (Hennessey et al., 2011). 'Familiar with the domain' is the conceptualization Amabile (1982) put forth in her original presentation of CAT, and still relies on (Hennessey et al., 2011). However, use of the word 'expert' is common in literature concerning CAT (e.g. Baer and McKool, 2009). 'Expert' has though been very broadly defined as it is dependent on product and domain how much knowledge is required, e.g. more knowledge is appropriate when deploying CAT in a more specialized domain such as physics or

computer programming (Hennessey, 1994). Moreover, Kaufman and Baer (2012) reported that when the purpose of a study is to investigate how ‘nonexperts’ assess creativity in some domain, their application could be acceptable. Here, the purpose is to assess creativity in relation to purchasability with a consumer product in mind, which could be assumed to fit that criterion well, as the focus is on appreciation of creativity.

With this requirement of CAT in mind, participants in the current study were experts, as they were required to be knowledgeable about the domain of products used in the experiment design. In modern Scandinavian society, the use of smartphones is on the rise, 45% of Danes and 50% of Norwegians own smartphones (IAB-Europe, 2012). It was assumed that smartphone users would be knowledgeable about amateur photography due to highly popular applications for photo sharing and picture quality amendments, as well as other accessories intended to enhance smartphone usage. Therefore it was concluded that smartphone accessories would be appropriate in the experiment. The age group (23-28 years) consequently has the highest smartphone usage rate of all age groups, or 62% (Go-Gulf 2012). The participant population consisted of master-level engineering students of mostly Scandinavian descendant. Collected responses were 22 in total, each assessing 15 products on the four attributes, producing an adequate dataset for further analysis conforming to CAT standards.

3.3. Stimuli

A web-based assessment was constructed where the study design was inspired by one described in Hennessey’s (1994) study. First, each participant was asked to get familiarized with the 15 products by reviewing them, in a randomized order, on the first slide without any assessing taking place. This was done to comply with the requirement of CAT, which is that products must be assessed relative to one another, thus forming a baseline by asking participants to first review all products in the assessment. On the second slide the products appeared again, in a random order, where participants were asked to review them in order to assess the creativity level for each product, on a seven point Likert scale. This procedure was repeated on a new slide for the remaining three attributes, technical advancement, aesthetic appeal and purchasability.

3.4. Procedure

The experiment was conducted in the beginning of a creativity course, taught at Technical University of Denmark, resulting in participant’s minimal knowledge of construct definitions and no introduction to the assessment technique, as this is required when conducting CAT (Amabile, 1982). Simple instructions were given in the beginning where participants were asked to review photos of the products and a short description text with each, to form a baseline, and consequently continue with the assessment. No predetermined criteria were provided in regards to creativity and its assessment, and assessment scales were numbered one to seven.

4. Results

For the purposes of establishing reliability results a Cronbach’s alpha calculation was performed for each of the rated attributes, as is standard procedure when utilizing CAT. Reliability proved to be acceptable for all four attributes, showing high interrater agreement. The acceptable correlation coefficient level is .7 (Hennessey et al., 2011), making results here highly acceptable: creativity level =.893, technical advancement =.939, aesthetic appeal =.907 and purchasability =.809.

The group average was run through a regression model for each object, where purchasability was the dependent variable regressed with the other three factors. Results from that analysis showed that purchasability can be predicted by the other three factors combined ($R^2=.927$), however, when the three factors were independently analyzed aesthetic appeal was only shown to predict purchasability with high beta (.896) and t values (9.932) as well as being significant at t value ($p=.000$). Technical advancement had low beta (.093) and t values (.937) and no significance ($p=.369$). Creativity further had a low beta level (.151), a low t value (1.664) and no significant relationship to purchasability ($p=.124$). Thus, the results revealed that H1 was not supported, as creativity level was not shown to

have a significant relationship ($p=.124$) with purchasability, i.e. creativity level of a product was not proven to have predictive validity in terms of purchasability.

Further analysis was performed on the statistical results of the experiment revealing first average ratings for each factor. The average rating creativity level received was $m=3.55$ out of the 7 point scale, average for technical advancement was $m=3.26$, aesthetic appeal received the highest average rating of $m=3.73$ and finally purchasability got $m=3.46$. Analysis was moreover performed on correlations between factors where some interesting results were identified, which are in line with results from the regression model. Pearson correlation calculations between creativity level and purchasability reveal correlations ranging from negative ($r=-.244$) to positive ($r=.773$) further supporting a weak or a non-existing relationship between creativity and purchasability. As aesthetic appeal was the only factor shown to predict purchasability ($p=.000$), it was of interest to look at the Pearson correlation between the two factors, which revealed commonly high positive correlations, ranging from $r=.308$ to $r=.782$, with eight of the items revealing over $r=.6$. Moreover, average ratings for each item are commonly highly similar for aesthetic appeal and purchasability, further supporting the existing relationship between the two factors. The relationship identified between aesthetic appeal and purchasability is an interesting side finding, and will be further discussed below.

5. Discussion

Horn and Salvendy (2006b; 2009) were first to establish a relationship between creativity and consumer behavior, i.e. high creativity levels were found to have predictive value in terms of willingness to purchase. That result is an important conclusion indicating that design creativity is a valuable attribute for organizations to strive for. The purpose of this study was to reproduce similar results when applying CAT as Horn and Salvendy were able to portray using the PCMI. However, the results of the current study did not identify design creativity as having a significant relationship to purchasability, in fact aesthetic appeal was the only product attribute that was shown to have predictive value for purchasability. That conclusion led the investigation towards looking more closely into the differences and/or similarities between the two constructs; creativity and aesthetic appeal, revealing that there are two different standpoints relevant to those discussions. One standpoint holds that aesthetic appeal is part of creativity and the other standpoint regards aesthetic appeal and creativity as being opposites. Horn and Salvendy actually take the first standpoint and hold that product creativity should incorporate aesthetic appeal as part of the core product attributes, and come to the conclusion that creativity drives purchasability. This is a common understanding within consumer behavior literature, however, it challenges the conceptual independence of the two constructs as creativity is defined as something novel and appropriate, while aesthetic appeal is usually associated with familiarity (Christensen et al., 2009). This leads to the second standpoint other scholars take who see aesthetic appeal and creativity as opposites and come to the conclusion that aesthetic appeal drives purchasability (Bloch, Brunel and Arnold, 2003).

Both standpoints have valid points to them and should be considered by designers. If leaning towards the first standpoint the argument would be that the two concepts should not be separated but rather thought of as deriving from the same processes, thus designers should strive towards a balance between novelty and familiarity. The other standpoint argues that the concepts are independent of each other. An important notion to point out is that both creativity and aesthetic appeal are highly subjective constructs and when being judged elicit emotional impact in participants (Christensen et al., 2009). Thus it could be assumed that designers should consider providing affect in product designs by integrating desirable and pleasing characteristics (Horn and Salvendy, 2009).

The results of the current study revealed that aesthetic appeal was the only product attribute to predict purchasability, thus seemingly fitting within the second standpoint. However the nature of the relationship between design creativity and aesthetic appeal was not investigated deterring further discussion of the relationship of the two constructs here. Our previous research has showed a significant relationship between aesthetic appeal and creativity, further supporting the complexity of this dilemma, which has been studied by several researchers (for a summary, see Christensen et al., 2009).

One might conclude from the results of the current study that CAT is not appropriate when determining attributes such as purchasability, and potentially other attributes that relate to consumer behavior. It is a highly appraised assessment method when it comes to creativity, and aesthetic appeal for that matter, as both concepts are very much subjective as the assessment method itself. But it might not be the right tool to measure willingness to buy, thus not sufficient when assessing the consequences of measured creativity, i.e. purchasability might simply not be a relevant factor when applying CAT. When looking at the statistical results purchasability got the lowest Cronbach Alpha score further supporting that conclusion.

But in the end it is important to keep in mind the reality that products designers face, which is that Horn and Salvendy (2006b; 2009) did establish a link between design creativity and consumer behavior and by portraying that link, revealing that design creativity, when recognized, is appreciated and is likely to make people want to acquire the creative product. Even though the current study could not come to the same conclusion when applying CAT, this is an important managerial implication for organizations competing on the product design market.

6. References

- Amabile, T. M. (1982). Social psychology of creativity: A consensual assessment technique. *Journal of Personality and Social Psychology*, 43 (5), 997-1013.
- Baer, J., & McKool, S. S. (2009). Assessing creativity using the consensual assessment technique. In C. Schreiner (Ed.), *Handbook of assessment technologies, methods, and applications in higher education* (pp. 65-77). Hershey, PA: IGI Global.
- Besemer, S. P., & Treffinger, D. J. (1981). Analysis of Creative Products: Review and Synthesis. *The Journal of Creative Behavior*, 15 (3), 158-178.
- Brown, D. C. (2008). Guiding computational design creativity research. In *Proceedings of the NSF International Workshop on Studying Design Creativity* (Vol. 8).
- Christensen, B. T., Kristensen, T., & Reber, R. (2009). Antecedents and consequences of creativity and beauty judgments in consumer products. *Creative Encounters Working Paper #29*.
- Go-Gulf. (2012, January 2012). *Blog*. Retrieved June 25, 2013, from Smartphone users around the world - statistics and facts: www.go-gulf.com/blog/smartphone
- Hennessey, B. A. (1994). The consensual assessment technique: An examination of the relationship between ratings of product and process creativity. *Creativity Research Journal*, 7 (2), 193-208.
- Hennessey, B. A., Amabile, T. M., & Mueller, J. S. (2011). Consensual Assessment. In M. Runco, & S. R. Pritzker, *Encyclopedia of Creativity*. Boston: Academic Press.
- Horn, D., & Salvendy, G. (2006a). Consumer-based assessment of product creativity: A review and reappraisal. *Human factors and ergonomics in manufacturing*, 16 (2), 155-175.
- Horn, D., & Salvendy, G. (2006b). Product creativity: conceptual model, measurement and characteristics. *Theoretical issues in ergonomic science*, 7 (4), 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 (3), 223-240.
- IAB-Europe. (2012). *Research*. Retrieved June 25, 2013, from Mediascope Europe - Denmark Launch presentation summary: http://www.iabeurope.eu/files/9913/7207/1264/MEDIASCOPE_2012_DENMARK_SUMMARY_LAUNCH_PRESENTATION.pdf
- Kaufman, J. C., & Baer, J. (2012). Beyond new and appropriate: Who decides what is creative? *Creativity Research Journal*, 24 (1), 83-91.
- Lau, S., Cheung, P. C., Lubart, T., Tong, T. M., & Chu, D. H. (2013). Bicultural effects on the creative potential of Chinese and French children. *Creativity Research Journal*, 25 (1), 109-118.
- Lebuda, I., & Karwowski, M. (2013). Tell me your name and I'll tell you how creative your work is: Author's name and gender as factors influencing assessment of products' creativity in four different domains. *Creativity Research Journal*, 25 (1), 137-142.
- Lu, C. C., & Luh, D. B. (2012). Comparison of assessment methods and raters in product creativity. *Creativity Research Journal*, 24 (4), 331-337.

- Mayer, R. E. (1999). Fifty years of creativity research. In R. J. Sternberg, *Handbook of Creativity* (pp. 449-460). Cambridge: Cambridge University Press.
- Runco, M. A., & Jaeger, G. J. (2012). The standard definition of creativity. *Creativity Research Journal*, 24 (1), 92-96.
- Sawyer, R. K. (2012). *Explaining Creativity: The Science of Human Innovation*. Second edition. New York: Oxford University Press.
- Stanko-Kaczmarek, M. (2012). The effect of intrinsic motivation on the affect and evaluation of the creative process among fine arts students. *Creativity Research Journal*, 24 (4), 304-310.
- Stein, M. I. (1953). Creativity and culture. *Journal of Psychology*, 36, 31-322.
- Sternberg, R. J. (1999). *Handbook of Creativity*. Cambridge: Cambridge University Press.
- Valgeirsdóttir, D., Onarheim, B. & Gabrielsen, G. (2014). Product Creativity Assessment of Innovations: Considering the Creative Process. *International Journal of Design Creativity and Innovation*: Under review.