

## **A STUDY ON STUDENT: ASSESSING FOUR CREATIVITY ASSESSMENT METHODS IN PRODUCT DESIGN**

**Yin, Yuan (1);  
Han, Ji (2);  
Huang, Shu (1);  
Zuo, Haoyu (1);  
Childs, Peter (1)**

1: Imperial College London;

2: University of Liverpool

### **ABSTRACT**

This paper asked participants to assess four selected expert-rated Taiwan International Student Design Competition (TISDC) products using four methods: Consensual Assessment Technique (CAT), Creative Product Semantic Scale (CPSS), Product Creativity Measurement Instrument (PCMI), and revised Creative Solution Diagnosis Scale (rCSDS). The results revealed that, between experts and non-experts, the ranking results by the CAT and CPSS were the same, while the ranking results of the rCSDS were different. The CAT, CPSS, and TISDC methods provided the same results indicating that raters may return the same results on creativity assessment, and the results are not affected by the selected methods. If it is necessary to use non-experts to assess creativity and the creativity results are expected to be the same with that of experts, asking non-expert raters to use CPSS to assess creativity and then ranking the creativity score is more reliable. The study offers a contribution to the creativity domain on deciding which methods may be more reliable from a comparison perspective.

**Keywords:** Creativity, Evaluation, Raters, Product, Conceptual design

### **Contact:**

YIN, Yuan  
Imperial College London  
Dyson School of Design Engineering  
United Kingdom  
y.yin19@ic.ac.uk

**Cite this article:** Yin, Y., Han, J., Huang, S., Zuo, H., Childs, P. (2021) 'A Study on Student: Assessing Four Creativity Assessment Methods in Product Design', in *Proceedings of the International Conference on Engineering Design (ICED21)*, Gothenburg, Sweden, 16-20 August 2021. DOI:10.1017/pds.2021.27

## 1 INTRODUCTION

Creativity is an important element in design, especially in product design (Chakrabarti, 2006). Designers expect to include creativity in a design to solve problems or satisfy customers' needs in new ways (Yilmaz and Seifert, 2011). The importance of creativity in design and prosperity in stimulating creativity methods has triggered the development of creativity assessment in design. However, the criteria for such creativity judgments are still controversial. *Novelty* is a common criterion in judgment (Dorst and Cross, 2001). The potential to be original and effective (Corazza, 2016), function (Acuna and Sosa, 2011) and appropriateness (Davis, 1986) may also be adopted as criteria.

The lack of clarity on which criteria to apply is associated with the ambiguous definition of creativity. Definitions of creativity can be divided into the three following conditions: ability to produce novelty (Guilford, 1950), ability to produce novelty and value (Sarkar and Chakrabarti, 2011; Jung et al., 2010), and ability to produce novelty and appropriateness (Chakrabarti, 2006; Fink et al., 2010; Davis, 1986). The different attributes of creativity will affect the decision on which criteria to apply; meanwhile, the differing criteria have been developed into different creativity assessment methods. This research aims to determine which existing creativity assessment methods in design are more reliable.

The main contributions of the paper are as follows:

1. This paper gives an evaluation of four creativity assessment methods to identify which method is more reliable for non-expert raters. It is also an attempt to practically compare the four methods and show at which levels the results of the four creativity assessment methods can be trusted.
2. The results of this paper indicate that, if ranking same-method-assessed products, expert-rated methods and non-expert-rated methods may provide the same results. This challenges the existing suggested raters for each creativity assessment method and experts may not be necessary.
3. This study can be a pilot study for researchers who are grappling with selecting a method that will be more reliable for them in recruiting participants to evaluate creativity in design.

## 2 LITERATURE REVIEW

Creativity assessment is mainly based on human actions at present (Lu and Luh, 2012). However, human judgment is subjective and/or error prone. Human-based creativity assessment usually involves people with various degrees of evaluation experience and different assessment criteria (Horn and Salvendy, 2006a). The existing assessment methods can be divided into subjective assessment, scoring, and structured method (Yen and Sun, 2008). Among the existing assessment methods, the Consensual Assessment Technique (CAT; Amabile, 1982) and Creative Product Semantic Scale (CPSS; Besemer and O'Quin, 1987) are usually the main creativity assessment methods applied (O'Quin and Besemer, 2006; White and Smith, 2001).

The CAT (Amabile, 1982; Horn and Salvendy, 2006b) relies on assessing final products. CAT does not have attributes or structures to follow. It is a scoring assessment method and only asks experts in a relevant domain to give a creativity score from 1 to 5 without explanation. Since it was first developed, CAT has been widely used and verified (Christiaans, 2002). The inter-rater reliabilities are usually more than 0.70, which is satisfactory (Baer, 2014; Amabile, 2018). However, the original CAT was promoted to evaluate at least two creativity outputs in a comparison. This limited its use because not all creativity outcomes are generated based on the same requirement, nor can they always be compared. However, this limitation has been solved. Baer et al. (2004) asked 13 experts to evaluate 103 stories, 103 personal narratives, and 102 poems written by eighth grade students. In their research, the three forms were compared. The results expanded the CAT from responding to the same assignments to comparing different assignments, especially in comparing which kinds of assignments have more creativity.

The CPSS (Besemer and O'Quin, 1987) is a further development of the Creative Product Analysis Matrix (CPAM). CPSS is a theoretical model that conceptualizes three dimensions: Novelty (original, surprising and germinal), Resolution (valuable, logical, understandable, and useful), and Elaboration and Synthesis (organic, elegant, complex, and well crafted) with a 7-point Likert scale. Although experts conduct the assessment and may bring some bias, the judgments of experts, non-experts, and naive users are the same. To be specific, the effects of experts, non-experts, and naive users on Novelty and Elaboration and Synthesis have been proven to have no influence, especially in terms of experts and naive users (O'Quin and Besemer, 1989). However, whether judgment on Resolution is

the same among the three groups of raters is not certain, while the evaluation of Elaboration and Synthesis is affected by Novelty and Resolution (Cropley and Urban, 2000). Similar results were reported by Besemer (1998), who evaluated three different chairs and found most coefficients to be in excess of 0.80 (reliabilities of the three dimensions ranging from 0.69 to 0.87). White et al. (2002) asked 43 professionals, 189 undergraduates, and 61 members of public to evaluate the advertisements in two booklets using a 15-item CPSS. The results showed that the evaluations were similar in the originality and logic judgments among professionals, undergraduates, and the public, but they were different for the well-crafted categories (Elaboration and Synthesis).

CAT and CPSS were both suggested to employ experts in design as raters (Horn and Salvendy, 2006a). This requirement brings a logistics challenge, especially when the assessment involves a large number of products (Baer, 1991). Therefore, some researchers have suggested that non-experts can be recruited to replace the experts (Niu and Sternberg, 2001). Some researchers tried to ask non-experts use CAT and CPSS to evaluate creativity. At the same time, new creativity assessment methods were also promoted. The Product Creativity Measurement Instrument (PCMI; Horn and Salvendy, 2009) and Creative Solution Diagnosis Scale (CSDS; Cropley and Cropley, 2008) were in response to this requirement. The two methods suggested recruiting non-experts as raters.

The PCMI employs a 7-point Likert scale (extremely not creative to extremely creative) to assess three main product creativity factors: Affect, Importance, and Novelty. It has a good inter-rater reliability coefficient (0.81; Horn and Salvendy, 2009). Notably, the PCMI is an assessment method in which non-experts have a higher inter-rater reliability than experts do (Lu and Luh, 2012). This was demonstrated by Lu and Luh (2012), who asked 9 experts and 11 non-experts to evaluate 70 works from the 2009 Taiwan International Student Design Competition (TISDC) based on the CAT and PCMI. The results showed that PCMI is effective in assessing product creativity, at least in design competitions. Both experts and non-experts have a high reliability and construct validity for assessing creativity. However, the results of non-experts have a higher inter-rater reliability compared with experts. Therefore, compared with experts, PCMI has a higher reliability and validity for non-experts. Specifically, the explanatory power for expert is 51%, while that of non-experts is 54.4%.

The CSDS has been suggested to be applied to humble products (Cropley and Cropley, 2008). In addition, it could be used by non-experts and single raters. For example, the CSDS can be employed by teachers, who are not experts in creativity but need to judge their students' creativity quickly and independently. The assessment is based on Relevance and Effectiveness, Generation of Novelty, Elegance, and Genesis. With the reliability of the scale being 0.79 (Cropley and Cropley, 2008), Cropley and Cropley (2008) expected that the CSDS could be universally applied to functional creativity evaluation among both experts and non-experts. To prove this, some research was conducted by Haller et al. (2011) with five experts and five non-experts evaluating students' work and no difference between the two groups' ratings was found.

It is understandable that the results of creativity from non-expert-rated methods will have a low correlation with that of expert-rated methods (Pétervári, 2018). Experts have a higher demand on creativity and individual creativity attitudes. Specifically, the grades of non-experts should be higher than those of experts. However, are the suggested raters (experts or non-experts) the only choice for a particular method? This triggers the question of whether the differences between the two kinds of raters are obvious enough to be identified as differences.

As mentioned above, most existing studies have focused on evaluating whether a particular method has a higher reliability in non-experts or experts. This has always been based on asking experts and non-experts to score some products using a single method. To extend this research, the present paper seeks to rank the scores instead of considering them independently to identify whether the results of ranking scores would be the same among experts and non-experts. In addition, apart from Lu and Luh (2012), little research has compared the reliability of the different creativity assessments with the types of raters. Therefore, this research also plans to compare the reliability of four creativity assessment methods in relation to the rater type.

### 3 METHOD

Details of the study for the data collection process are presented in this section, focusing on the participants, study design, and procedure.

### 3.1 Participants

This creativity assessment evaluation included 62 Chinese people (32 females and 30 males) with product design, graphic design, visual media, and industrial design backgrounds. The participants were between 19 and 35 years old. All participants were provided with an information sheet about the study and signed a consent form to take part in the study. The whole study was conducted based on an online questionnaire by Qualtrics. Nearly all participants finished the study in 50 minutes.

Experts and non-experts can both be raters. The paper selected more easily accessible raters—non-experts—as the raters. The results for experts were directly indicated from the selection of products. The definition on experts in design is various. For example, a designer with 2 years of experience may define him/herself as expert, but a designer with 3 years of experience might define him/herself as non-expert. Thus, participants were asked to self-define if they are non-experts in this study. This self-definition reflected their creativity evaluation experience. In this study, we first asked the participant to answer if they think they are an expert or non-expert. Then, they need to answer a question on how many years they have been involved in design. Then, they were asked to report how many creativity assessments they have done. There were two participants who defined themselves as experts; they were excluded from this study. All of participants were in design less than 6 years and had evaluated fewer than 50 creativity outputs.

### 3.2 Study design

Participants needed to complete an online questionnaire in under 60 minutes. The questionnaire collected information about participants' educational background and their creativity evaluation experience and asked participants to evaluate the creativity of four selected products by the CAT, CPSS, CSDS, and PCMI creativity assessment methods.

#### 3.2.1 Criteria selection

The creativity assessment methods were limited to those that had been developed for more than 10 years, the latest version having been evaluated by someone other than the developers, and the assessment being based on ranking or having structures to assess. With these requirements, four methods were finally selected—the CAT, CPSS, PCMI, and CSDS. The reason why the selection criteria include “the methods had been developed for more than 10 years” is that the methods promoted more than 10 years have more chance to be a mature assessment method and be verified by other researchers (Chien et al., 2020); “the latest version should have been evaluated by someone other than the developers” was included as criteria to ensure the methods have been proved or supported by some more researchers; the selection criteria including “the assessment should be based on ranking or having structures to assess” is because experts raters or non-experts raters often need to assess creativity based on these two kinds of assessment forms (Yen and Sun, 2008).

Both expert-rated and non-expert-rated methods were considered. To be specific, although the CPSS and CAT suggest employing experts as raters, the paper still included the two methods. This is because the two methods can also be used by non-experts, and there is no research mentioning that non-expert raters' CAT or CPSS results are less reliable than their assessment results of using the non-expert-rated methods (Haller et al., 2011).

The CAT asks raters to give 1–5 creativity points without explanation. The CPSS includes 3 categories, 11 subscales, and 55 items. However, we only selected the 18 items represented in product design. The selected items are based on the suggestion of Han (2018) and Chulvi et al. (2012). The PCMI uses a 7-point Likert scale with 13 items to assess three main product creativity factors: Affect, Importance, and Novelty. CSDS previously included 30 items based on four criteria. However, after testing the six items (Operability, Durability, Replication, Incrementation, Reconstruction, and Recognition), the test showed that the cross loading is  $> 0.3$  or loaded onto no factors (Kaufman and Baer, 2012). Therefore, the revised CSDS (rCSDS), which uses 24 items with a 3-point Likert scale, was selected in this study.

The assessment objects can be daily activities, design activities, products, and divergent thinking outputs. Daily activities and design activities are more likely to be used as methods to evaluate human creativity instead of a product. Divergent thinking outputs can be a quick and brief reflection on design and may not reflect a product's creativity accuracy. Therefore, in this study, products were selected as the assessment object.

Products in the 2019 TISDC were selected as the source of expert-rated results. The 2019 TISDC recruited Chinese experts with a product design background to rank the works with Gold, Silver, Bronze, and Honorable Mention Prizes. This provides an “indication” of which design is better. In addition, this competition provides a quantitative evaluation standard, which is “Creativity: 50%; Thematic: 30%; Design Presentation: 20%.” Since creativity, which is the main evaluation criterion in the prize, is the focus for this study, this competition was deemed a good choice by the authors. The four products awarded Gold, Silver, Bronze, and Honorable Mention were selected and named Product 1, Product 2, Product 3, and Product 4. The ranking result for creativity was that Product 1 was higher than Product 2, Product 2 was higher than Product 3, and Product 3 was higher than Product 4.

### 3.3 Procedure

After potential participants signed the digital consent form, the researchers sent them online questionnaire links via email. Generally, participants finished the questionnaire in one hour. The questionnaire is divided into two parts. The first part is designed to collect the participants’ creativity assessment experience data. Then, the four products and the four methods were matched with each other. The sixteen assessments were shown in the questionnaires at random order. For each assessment, participants were first given the product poster. Then the introduction about how to use the pointed assessment method to assess the product and how long they were suggested to use was given.

There is no time limitation on each assessment method. For CAT assessment, participants were informed that they need to give a creativity point to the product from 1 to 5 without any justification; for CPSS and PCMI assessment, participants were asked to assess the product by finishing a list based on the bipolar 7-point scale; for rCSDS, participants were asked to assess the product by finishing the list based on a 3-point scale of words.

## 4 RESULTS

The results from the questionnaires are analyzed in this section. This part includes four subsections: the results from mean value analysis, the results from standard deviation, the results from cross-graph analysis and the principal findings. The extended findings are discussed in “Discussion” session. Two terms used in the followings need to be defined first - “each-method-assessed” and “same-method-assessed”. “Each-method-assessed” products refer to the single product assessed by a single method. “Same-method-assessed” products mean the four products assessed by the same method.

*Table 1. The mean creativity score of each-method-assessed product, standard deviation of each-method-assessed product, and the order of same-method-assessed products’ mean value*

| Method | Product number | Maximum score that the method can obtain | Mean   | Std. Deviation | the order of same-method-assessed products’ mean creativity score |
|--------|----------------|--|--------|----------------|---|
| CAT    | Product1       | 5  | 4.20   | .632           | 1   |
|        | Product2       |  | 3.95   | .769           | 2   |
|        | Product3       |  | 3.58   | .907           | 3   |
|        | Product4       |  | 3.45   | .769           | 4   |
| CPSS   | Product1       | 126                                      | 101.63 | 12.453         | 1   |
|        | Product2       |  | 95.37  | 17.920         | 2   |
|        | Product3       |  | 94.18  | 18.287         | 3   |
|        | Product4       |  | 88.48  | 14.860         | 4   |
| PCMI   | Product1       | 98                                       | 61.53  | 13.702         | 1   |
|        | Product2       |  | 59.15  | 13.756         | 2   |
|        | Product3       |  | 57.77  | 15.010         | 3   |
|        | Product4       |  | 55.08  | 13.066         | 4   |
| rCSDS  | Product1       | 120                                      | 74.87  | 9.525          | 1   |
|        | Product2       |  | 71.75  | 11.813         | 3   |
|        | Product3       |  | 72.50  | 13.021         | 2   |
|        | Product4       |  | 71.90  | 8.457          | 4   |

#### 4.1 Mean creativity scores of each-method-assessed products

First, the mean creativity scores of each-method-assessed products were calculated. The same-method-assessed products' mean creativity scores were then ranked. This approach was used to identify whether the order of same-method-assessed products' mean creativity scores would be consistent with those of the TISDC. The mean value is shown in Table 1. When participants used the CAT (maximum score is 5) to assess the four given products, the average creativity scores were 4.20, 3.95, 3.64, and 3.58 for Products 1 to 4 respectively. The order of these average creativity scores is consistent with the TISDC results. Similar results were also generated when participants used the CPSS (maximum score is 126) and PCMI (maximum score is 98) to rate the four products. The average creativity scores of CPSS-assessed products were 101.63, 95.37, 94.18, and 88.48, while those of PCMI-assessed products were 61.53, 59.15, 57.77, and 55.08. These average score results suggest that the CAT, CPSS, and PCMI have the potential to allow non-experts to evaluate the product with a certain accuracy.

Different results from the CAT, CPSS, and PCMI are obtained with the rCSDS approach. When participants used the rCSDS (maximum score is 120) to assess the four given products, the average creativity score ratings were 74.87, 71.75, 72.50, and 71.90 for Products 1 to 4.

#### 4.2 Standard deviation of each-method-assessed products

To quantify the dispersion of the data values, standard deviation was introduced. The standard deviation of each-method-assessed product is also shown in Table 1. The standard deviations of the CAT-assessed products were 0.632, 0.769, 0.769, and 0.907. Those of the CPSS-assessed products are 12.453, 17.920, 18.287, and 14.860. The results for the PCMI-assessed products were 13.702, 13.756, 15.010, and 13.066. Finally, the standard deviations of the rCSDS-assessed products were 9.525, 11.813, 13.021, and 8.457.

#### 4.3 Cross-graph analysis

Although the order of the CAT-, CPSS-, and PCMI-method-assessed products' mean creativity scores was consistent with that of the TISDC, and the results of the rCSDS were different from that of the TISDC, statistical analysis was required to determine whether there were significant differences between the order of same-method-assessed products and the order of TISDC. This statistical analysis relied on SPSS. Since each participant's same-method-assessed product creativity score was considered an ordinal variable, cross-graph analysis was selected to test whether there was a significant difference between the orders of the same-method-assessed products and the order of the TISDC. The analysis of each method included the Person chi-square, contingency-coefficient (C), gamma value (G), gamma sig, and Kendall's tau-b (Tb) tests.

The Person chi-square test was conducted to analyze whether the method and the order of method-assessed products were independent. In the test results, if a p-value was less than 0.05, it indicated that the two data sets were related. The C test was conducted to support the related levels. The gamma value and gamma sig tests were conducted to identify whether the order of method-assessed products was related to the order in the TISDC. If the gamma sig was less than 0.05, it indicated that the results were statistically significant. Kendall's tau-b test was conducted as an additional supportive test of whether the order of method-assessed products was related to the order in the TISDC. For the C value, the gamma value tests, and the Tb value, the higher the absolute value, the higher the relation; a positive number indicated a positive correlation, while a negative number indicated a negative correlation.

For the CAT creativity assessment method, the method and the order of method-assessed products were not significantly independent ( $p=.000$ ,  $C=.355$ ). The order of method-assessed products was significantly negatively related to the order in the TISDC ( $G=-.414$ ,  $p=.000$ ). This negative relation was supported by Tb ( $Tb= -.297$ ). In the study, the higher the method-assessed product number, the lower creativity the product had. Therefore, the results showed that the higher the ranking results in the TISDC indicated, the smaller the method-assessed product number became. In other words, the results of the CAT were consistent with those of the TISDC.

For the CPSS creativity assessment method, the method and the order of method-assessed products were independent ( $p=.319$ ,  $C=.672$ ). The order of method-assessed products was significantly negatively related to the order in the TISDC ( $G=-.242$ ,  $p=.000$ ). This negative relation was supported by Tb ( $Tb= -.207$ ). In other words, the results of CPSS were consistent with those of the TISDC.

For the PCMI creativity assessment method, the method and the order of method-assessed products were independent ( $p=.295$ ,  $C=.642$ ). The order of method-assessed products was negatively related to the order in the TISDC, but the result was not significant ( $G=-.135$ ,  $p=.010$ ). This negative relation was supported by  $T_b$  ( $T_b = -.116$ ). In other words, the results of PCMI were not consistent with those of the TISDC.

For the rCSDS creativity assessment method, the method and the order of method-assessed products were independent ( $p=.807$ ,  $C=.579$ ). The order of method-assessed products was negatively related to the order in the TISDC, but the result was not significant ( $G=-.074$ ,  $p=.154$ ). This negative relation was supported by  $T_b$  ( $T_b = -.063$ ). In other words, the results of rCSDS were not consistent with those of the TISDC.

#### 4.4 Principal Findings

From the two analyses, the paper can conclude that from meaning value analysis, non-experts and experts can get same creativity ranking results when using CAT, CPSS or PCMI. From cross graphic analysis, the results of CAT and CPSS are statistically significant negative relation with that of TISDC. Based on the two results, non-experts and experts can get reliable same creativity ranking results when using CAT and CPSS. Creativity score of the product is not independent of CAT. In other words, the creativity score is affected by CAT method. Creativity score of the product is independent of CPSS. In other words, the creativity score is not affected by CPSS method. Therefore, if products need non-experts to assess creativity and the creativity results are expected to be the same with that of experts. The paper suggests asking non-expert raters use CPSS to assess creativity and then ranking the creativity score. Based on the findings in this paper, this approach could make the results of non-expert-raters the same as that of experts and increase the reliability of the results.

### 5 DISCUSSION

The results from the gamma value are not consistent with those from the mean values of the CAT, CPSS, PCMI, and rCSDS. When ranking the same-method-assessed products, the mean value results of the CAT, CPSS, and PCMI are the same as those of the TISDC, while the results of the rCSDS are different. However, when using cross-graph analysis, the results show that the results of the CAT and CPSS are the same as those of the TISDC, while the results of the PCMI and rCSDS are different. The reason why the results in rCSDS were different with others may be that although generally, CAT, CPSS, PCMI and rCSDS were used to assess product creativity, rCSDS is more applicable in assessing humble products (Cropley and Cropley, 2008). The selected four products may not be good examples of humble products and therefore, the results of rCSDS could not assess creativity of these products accurately.

When considering mean values, there is an interesting difference. The PCMI and rCSDS are both methods used by non-experts to evaluate creativity. Although the TISDC asked experts to evaluate the four given products, and the experts' results are considered in this paper, the participants using the PCMI assessment method could obtain the same results as the TISDC, while the rCSDS did not. In addition, although participants who are non-experts used the CAT and CPSS, which are suggested to be used by experts as raters, the results were the same as the TISDC, which had experts as raters. On the one hand, these results suggest that, although the CAT, CPSS, and PCMI were suggested to be used by non-experts or experts, the results of non-experts and experts may be the same when ranking creativity scores. On the other, the different results in the rCSDS may refute this view.

The cross-graph analysis seems to support that the PCMI and rCSDS, which are both methods used by non-experts to evaluate creativity, will provide different results from the CAT, CPSS, and TISDC, which suggests that experts should be included as raters. Contrasting results between the cross-graph and mean value are evident for the PCMI. The results from the mean value and cross-graph were different. However, this could be predicted in the standard deviation. The standard deviation in the PCMI is the highest among the four methods based on the maximum total score that each method can achieve. The higher the standard deviation is, the less concentrated the data are. The cross-graph is a test to evaluate whether results have a strong relationship. With less concentrated data, the results can be less reliable. However, since the p-value is 0.010, it may still be possible to show a significant relation. Therefore, if more samples are used, the results will be expected to show whether the mean or results of the cross-graph will be consistent. Although different results exist in relation to the PCMI compared with the other

approaches, this difference cannot overturn the possibility that, when ranking the creativity scores of the CAT, CPSS, and PCMI, non-experts and experts may have the same assessment results.

The ordering results were the same among the CAT, CPSS, and TISDC may suggest that the creativity assessment results from experts may be the same and not affected by the methods. This may be a reasonable conclusion. Although the definition of creativity in a product is different, the creativity for a particular product in a particular individual is the same. Therefore, it may be predicted that, if the methods are used to assess creativity in a product, the results have the possibility of being the same.

Although the four products were selected from the TISDC, and the ordering was given by experts, it is only a ranking result rather than a creativity score. In addition, experts' definition of creativity is uncertain. Although the TISDC was mainly based on the "creativity levels" to assess products, other criteria were also included. These could be limitations of this study. In other words, it might be better to conduct an expert evaluation to confirm the creativity ranking instead of directly using the results from TISDC because Thematic and Design Presentation also affect the ranking. Furthermore, participants have used all four methods, so it might affect the actual use of the assessment methods even the random order has been applied. For example, raters used CAT as the first assessment. The CAT evaluation result might affect the participants' evaluation performance on using CPSS. In addition, participants may previously have had accessed to these selected products or have an impression on the results of TISDC prizes. Such access and impressions may affect the results of assessment in this study. Any future study needs to consider the influence of whether participants have seen the assessed product or the prize of the assessed product before. Another limitation is that the study was only applied to 60 Chinese participants; nevertheless, the results could potentially be used as a pilot study to challenge the existing creativity assessment suggestion about who can serve as raters. If the ranking results of creativity assessment are the same between experts and non-experts, the existing method-suggested raters may become less reliable. A further study involving more samples, more products, and raters' experience levels is planned to produce additional insights.

## 6 CONCLUSION

In this paper, a questionnaire was administered to assess the creativity of four selected products. The study results indicated that, when using the CAT and CPSS, the ranking results were the same between experts and non-expert raters, while the ranking results of the rCSDS were different between experts and non-expert raters. More research needs to be done to identify whether the PCMI's ranking creativity results will be the same between experts and non-expert raters. If products need non-experts to assess creativity and the creativity results are expected to be the same with that of experts, the findings from this study suggests asking non-expert raters use CPSS to assess creativity and then ranking the creativity score. The results may also help researchers and designers who want to assess creativity by providing comparison results of the four creativity methods. This will give researchers and designers a direct view on which methods may be more reliable. Furthermore, the study indicated a new way to evaluate the four creativity assessment methods.

## REFERENCES

- Acuna, A. and Sosa, R. (2011), "The complementary role of representations in design creativity: Sketches and models", *Design creativity 2010*, Springer, London, pp. 265-270. [https://doi.org/10.1007/978-0-85729-224-7\\_34](https://doi.org/10.1007/978-0-85729-224-7_34)
- Amabile, T. M. (2018), *Creativity in context: Update to the social psychology of creativity*, Routledge.
- Baer, J. (1991), "Generality of creativity across performance domains", *Creativity Research Journal*, Vol. 4 No. 1, pp. 23-39. <https://doi.org/10.1080/10400419109534371>
- Baer, J. (2014), "Creativity and divergent thinking: A task-specific approach", Psychology Press.
- Baer, J., Kaufman, J. C., and Gentile, C. A. (2004), "Extension of the consensual assessment technique to nonparallel creative products", *Creativity research journal*, Vol. 16 No. 1, pp. 113-117. [https://doi.org/10.1207/s15326934crj1601\\_11](https://doi.org/10.1207/s15326934crj1601_11)
- Besemer, S. P. (1998), "Creative product analysis matrix: testing the model structure and a comparison among products--three novel chairs", *Creativity Research Journal*, Vol. 11 No. 4, pp.333-346. [https://doi.org/10.1207/s15326934crj1104\\_7](https://doi.org/10.1207/s15326934crj1104_7)
- Besemer, S. P. and O'Quin, K. (1987), "Creative product analysis: Testing a model by developing a judging instrument", *Frontiers of creativity research: Beyond the basics*, pp. 367-389.



- Chakrabarti, A. (2006), "Defining and supporting design creativity", *DS 36: Proceedings DESIGN 2006, the 9th International Design Conference*, Dubrovnik, Croatia, pp. 479-486.
- Chien, Y. C., Liu, M. C., and Wu, T. T. (2020), "Discussion-record-based prediction model for creativity education using clustering methods", *Thinking Skills and Creativity*, Vol.36, pp. 100650. <https://doi.org/10.1016/j.tsc.2020.100650>
- Christiaans, H. H. (2002), "Creativity as a design criterion", *Communication Research Journal*, Vol. 14 No. 1, pp. 41-54. [https://doi.org/10.1207/s15326934crj1401\\_4](https://doi.org/10.1207/s15326934crj1401_4)
- Corazza, G. E. (2016), "Potential originality and effectiveness: The dynamic definition of creativity", *Creativity research journal*, Vol. 28 No. 3, pp. 258-267. <https://doi.org/10.1080/10400419.2016.1195627>
- Cropley, A. J., and Urban, K. (2000), "Programs and strategies for nurturing creativity", *International handbook of giftedness and talent*, pp. 485-498. <https://doi.org/10.1016/b978-008043796-5/50034-6>
- Cropley, D. and Cropley, A. (2008), "Elements of a universal aesthetic of creativity", *Psychology of Aesthetics, Creativity, and the Arts*, Vol. 2 No. 3, pp. 155-161. <https://doi.org/10.1037/1931-3896.2.3.155>
- Davis, G. A. (1986), *Creativity is forever*, Kendall/Hunt Publishing, Dubuque.
- Dorst, K. and Cross, N. (2001), "Creativity in the design process: co-evolution of problem-solution", *Design studies*, Vol. 22 No. 4, pp. 425-437. [https://doi.org/10.1016/s0142-694x\(01\)00009-6](https://doi.org/10.1016/s0142-694x(01)00009-6)
- Fink, A., Grabner, R. H., Gebauer, D., Reishofer, G., Koschutnig, K., and Ebner, F. (2010), "Enhancing creativity by means of cognitive stimulation: Evidence from an fMRI study", *NeuroImage*, Vol. 4, pp. 1687-1695. <https://doi.org/10.3389/conf.fnins.2010.11.00025>
- Guilford, J. (1950), "Creativity", *American psychologist*, Vol. 5, pp. 444-454.
- Haller, C. S., Courvoisier, D. S., and Cropley, D. H. (2011), "Perhaps there is accounting for taste: Evaluating the creativity of products", *Creativity Research Journal*, Vol. 23 No. 2, pp. 99-109. <https://doi.org/10.1080/10400419.2011.571182>
- Han, J. (2018), *Combinational creativity and computational creativity*, Doctoral dissertation, Imperial College London.
- Horn, D. and Salvendy, G. (2006a), "Consumer-based assessment of product creativity: A review and reappraisal", *Human factors and ergonomics in manufacturing and service industries*, Vol. 16 No. 2, pp. 155-175. <https://doi.org/10.1002/hfm.20047>
- Horn, D., and Salvendy, G. (2006b), "Product creativity: conceptual model, measurement and characteristics", *Theoretical Issues in Ergonomics Science*, Vol. 7 No. 4, pp. 395-412. <https://doi.org/10.1080/14639220500078195>
- Horn, D., and Salvendy, G. (2009), "Measuring consumer perception of product creativity: Impact on satisfaction and purchasability", *Human Factors and Ergonomics in Manufacturing and Service Industries*, Vol. 19 No. 3, pp. 223-240. <https://doi.org/10.1002/hfm.20150>
- Jung, R. E., Segall, J. M., Jeremy Bockholt, H., Flores, R. A., Smith, S. M., Chavez, R. S., and Haier, R. J. (2010), "Neuroanatomy of creativity", *Human brain mapping*, Vol. 31 No. 3, pp. 398-409. <https://doi.org/10.1002/hbm.20874>
- Kaufman, J. C., and Baer, J. (2012), "Beyond new and appropriate: Who decides what is creative?", *Creativity Research Journal*, Vol. 24 No. 1, pp. 83-91. <https://doi.org/10.1080/10400419.2012.649237>
- Lu, C. C., and Luh, D. B. (2012), "A comparison of assessment methods and raters in product creativity", *Creativity research journal*, Vol. 24 No. 4, pp. 331-337. <https://doi.org/10.1080/10400419.2012.730327>
- Niu, W., and Sternberg, R. J. (2001), "Cultural influences on artistic creativity and its evaluation", *International journal of psychology*, Vol. 36 No. 4, pp. 225-241. <https://doi.org/10.1080/00207590143000036>
- O'Quin, K., and Besemer, S. P. (1989), "The development, reliability, and validity of the revised creative product semantic scale", *Creativity Research Journal*, Vol. 2 No. 4, pp. 267-278.
- O'Quin, K., and Besemer, S. P. (2006), "Using the creative product semantic scale as a metric for results-oriented business", *Creativity and Innovation Management*, Vol. 15 No. 1, pp. 34-44. <https://doi.org/10.1111/j.1467-8691.2006.00367.x>
- Petervari, J. (2018), *The evaluation of creative ideas—analysing the differences between expert and novice judges*, Doctoral dissertation, Queen Mary University of London.
- Sarkar, P., and Chakrabarti, A. (2011), "Assessing design creativity", *Design studies*, Vol. 32 No. 4, pp. 348-383. <https://doi.org/10.1016/j.destud.2011.01.002>
- White, A., Shen, F., and Smith, B. L. (2002), "Judging advertising creativity using the creative product semantic scale", *The Journal of Creative Behavior*, Vol. 36 No. 4, pp. 241-253. <https://doi.org/10.1002/j.2162-6057.2002.tb01067.x>
- White, A. and Smith, B. L. (2001), "Assessing advertising creativity using the creative product semantic scale", *Journal of Advertising Research*, Vol. 41 No. 6, pp. 27-34. <https://doi.org/10.2501/jar-41-6-27-34>
- Yen, J. and Sun, C. (2008), "A study on the design teachers' and students' recognition of performance evaluation and evaluation Criteria", *Journal of Science and Technology*, Vol. 17 No. 1, pp. 41-57.
- Yilmaz, S., and Seifert, C. M. (2011), "Creativity through design heuristics: A case study of expert product design", *Design Studies*, Vol. 32 No. 4, pp. 384-415. <https://doi.org/10.1016/j.destud.2011.01.003>



CAMBRIDGE  
UNIVERSITY PRESS