

Central Lancashire Online Knowledge (CLoK)

Title	Design creativity technical execution and aesthetic appeal: A CAT with caveats (Part 2)
Type	Article
URL	https://clock.uclan.ac.uk/19958/
DOI	https://doi.org/10.1080/21650349.2017.1381043
Date	2017
Citation	Jeffries, Karl K., Zamenopoulos, T. and Green, A.J.K (2017) Design creativity technical execution and aesthetic appeal: A CAT with caveats (Part 2). International Journal of Design Creativity and Innovation, 6 (1-2). pp. 66-79. ISSN 2165-0349
Creators	Jeffries, Karl K., Zamenopoulos, T. and Green, A.J.K

It is advisable to refer to the publisher's version if you intend to cite from the work.
<https://doi.org/10.1080/21650349.2017.1381043>

For information about Research at UCLan please go to <http://www.uclan.ac.uk/research/>

All outputs in CLoK are protected by Intellectual Property Rights law, including Copyright law. Copyright, IPR and Moral Rights for the works on this site are retained by the individual authors and/or other copyright owners. Terms and conditions for use of this material are defined in the <http://clock.uclan.ac.uk/policies/>



Design Creativity, Technical Execution and Aesthetic Appeal: A CAT with Caveats (Part 2)

Journal:	<i>International Journal of Design Creativity and Innovation</i>
Manuscript ID	TDCI-2017-0007.R2
Manuscript Type:	Original Article
Keywords:	concept generation, creativity evaluation, creativity education, idea generation

SCHOLARONE™
Manuscripts

Pre-View Only

Design Creativity, Technical Execution and Aesthetic Appeal: A CAT with Caveats (Part 2)

This study explores to what extent technical execution and aesthetic appeal may be related to assessments of graphic design creativity. These new research findings build upon Jeffries' 2015 publication in the International Journal of Design Creativity and Innovation, and further underpin the caveats identified in relation to the Consensual Assessment Technique (CAT). Eight professional graphic designers rated thirty-two artworks for a creative typographical task. Individual artworks were created by novices who had no experience of graphic design, through to professional graphic designers with 35 years of full-time experience. Written instructions to judges emphasised artwork be rated on creativity only (without considering technical execution or aesthetic appeal), and this "creativity only" feature was verbally re-emphasised to judges by the researcher. Inter-rater agreement for creativity was a Cronbach's alpha of 0.92; considerably higher than in previous studies, with implications that may relate to the use of the CAT as a measure of design creativity more broadly, and beyond graphic design.

Keywords: Consensual Assessment Technique (CAT), graphic design, design creativity assessment

1. Introduction

"...A product or response is creative to the extent that appropriate observers independently agree it is creative. Appropriate observers are those familiar with the domain in which the product was created or the response articulated." (Amabile, 1982, p. 1001).

Amabile's description of creativity measurement through social consensus by domain experts has formed the theoretical framework for thousands of research projects. In turn, the Consensual Assessment Technique (CAT) has been shown by many researchers to offer a reliable and valid operational definition of creativity (Baer & McKool, 2009). In

1
2
3 its practical application, CAT guidance states that when creativity is assessed in a new
4 domain (i.e. one that has not been studied with a particular task), researchers should ask
5 judges to rate additional constructs, specifically, technical execution and aesthetic
6 appeal (Amabile, 1982; Hennessey, 1994), and check to see if creativity ratings are
7 distinct from these criteria.
8
9

10
11 For example, artworks in graphic design can vary in their level of technical refinement.
12
13 At one end of the spectrum are conceptual sketches, where the seed of an idea can be
14 perceived: even if the sketches lack refinement in, for example, font selection, layout or
15 composition; the creativity of the idea can still be evaluated. At the other end of a
16 technical spectrum are finished artworks; artworks that are ready to go to print or
17 publication, where every aspect of visual communication has been crafted and refined to
18 perfection by the designer.
19
20
21
22
23
24
25
26
27
28
29

30
31 In regard to CAT protocol, once the relationship between creativity, technical execution
32 and aesthetic appeal has been shown to be distinct for a given task, researchers need
33 only ask for ratings of creativity for subsequent studies and can assume that technical
34 execution and aesthetic appeal are no longer a conflating issue for validity. Indeed, in
35 Amabile's 1982 work the extent to which creativity may be isolated from such factors
36 was a formative part of her paper, and she concluded that "...although judges were not
37 provided with a definition of creativity...they consistently and reliably identified a
38 quality in both types of product that was distinct from technical execution. Moreover,
39 for artworks, it was distinct from aesthetic appeal as well" (p.1010).
40
41
42
43
44
45
46
47
48
49
50

51
52 Hennessey's discussion of the CAT (1994), however, acknowledged that for "some
53 domains" (p.195) the distinction between technical execution and aesthetic appeal may
54 be less clear and that creativity is likely to correlate with these aspects. Indeed, in
55
56
57
58
59
60

1
2
3 several design related studies evidence can be found of this. For example, Christiaans'
4 (2002) applied the CAT to Industrial Design Engineering students, and found strong
5 correlations between creativity and aesthetic criteria, but not for creativity and technical
6 quality. Wojtczuk and Bonnardel's (2012) use of the CAT found a positive correlation
7 between creativity and four other criteria (aesthetics, originality, brief appropriateness
8 and audience appropriateness) for a poster design task.
9
10

11
12 This study explores to what extent technical execution and aesthetic appeal may be
13 related to assessments of graphic design creativity. The results are relevant to graphic
14 design but also extend to other areas of design creativity, and consideration of CAT
15 protocol more broadly.
16
17

18 19 20 21 22 23 24 25 26 27 28 **2. Background**

29
30 Within the Creative Industries graphic design links across many sectors: be it the need
31 for marketing material; a new logo for an organisation; the presentation of scientific
32 information, or the development of a new product. Graphic design will play a part:
33 sometimes in the background, at other times centre stage.
34
35

36
37 Established within traditional print media alongside the development of the Digital
38 Creative Industries, the roots of graphic design, arguably, do not first begin in the 20th
39 century. Visual communication: turning words into icons, or using images *to say a*
40 *thousand words* is ancient. Like many established professions its boundaries blur and
41 change, and the line between graphic design, illustration, advertising or copywriting is
42 debatable and in flux (we could add other domains to this list: typography, printing, web
43 design, photography, animation, packaging, let alone new technologies searching to
44 establish an identity).
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

1
2
3 Yet, within this flux, graphic design exists and is global. Throughout the world,
4
5 universities, colleges and private companies run graphic design courses; their students'
6
7 gain qualifications that enable them to work for design agencies or set up their own
8
9 business, and they in turn work for clients who require graphic design to appeal to their
10
11 audience and customers. In this context, creativity is an asset and gives clients, design
12
13 agencies, and individual graphic designers an edge in a competitive market.
14
15

16
17
18 Since the first systematic review of the CAT in design publications in 2012 (Jeffries),
19
20 CAT has continued to follow an upward trend in growth within scholarly journals.
21
22

23 Indeed, more CAT studies and citations occurred in the five-year period between 2010-
24
25 2015 than in the previous 28 years of CAT research. The impact is that a greater number
26
27 of researchers, both within and outside of design, apply CAT as their operational
28
29 definition of creativity.
30
31

32
33
34 However, despite Amabile's original guidance, and subsequent guidance by Kaufman,
35
36 Plucker, and Baer (2008) who dedicated a complete chapter of their book *Essentials of*
37
38 *Creativity Assessment* to the CAT, as did Hennessey, Amabile, and Mueller (2011) for
39
40 the *Encyclopaedia of Creativity*, the way different researchers implement the CAT in
41
42 practice shows substantial inconsistencies (Cseh, 2014). To what extent differing
43
44 interpretations of the CAT protocol influence the reliability and validity of the method
45
46 remains a topic for international debate, and this paper continues to explore such
47
48 considerations in relation to design creativity, and graphic design specifically.
49
50
51

52
53
54 Prior to 2015, data from the use of CAT in graphic design were not widely published.

55
56 Once these data became available, it was clear that a graphic design CAT appeared
57
58
59
60

1
2
3 more sensitive to methodological protocols than previous studies in other domains had
4 reported. Access to these results served as an encouragement to understand further what
5 may be influencing these low levels of consensus and to revisit past assumptions upon
6 which the CAT was established. This paper explores three interrelated assumptions
7 around the effect of “creativity only” instructions, the level of correlation between
8 creativity, technical execution and aesthetic appeal, and the diversity of artwork
9 presented to judges.
10
11
12
13
14
15
16
17
18
19

20
21 CAT ratings are based on the normative assessment of creativity within a defined group
22 (i.e. the specific participants in a study). As a result, Jeffries (2015) argued that diversity
23 of artwork may play a significant factor in levels of consensus amongst judges, and
24 researchers should not assume that there is enough diversity in a sample of artwork
25 (particularly if the consensus is low).
26
27
28
29
30
31
32
33

34 How such diversity of artwork is achieved can occur in several ways. For example,
35 Jeffries’ 2015 study used pre-selection of artwork based on academic grades; other
36 researchers, like Christiaans and Venselaar (2005), randomly sampled from a
37 population of 240 students to select 55 and then used higher and lower creativity scores
38 to select 18 cases for qualitative analysis.
39
40
41
42
43
44
45
46

47 As an alternative to randomization or pre-selection, in the present study, participants
48 were stratified according to levels of expertise (novice, intermediate and expert) within
49 graphic design. The logic for stratification was based on the assumption that experts are
50 likely to produce more creative graphic designs than intermediates, and intermediates
51 more than novices. While in specific instances that assumption may not hold, it is the
52
53
54
55
56
57
58
59
60

1
2
3 group as a whole that forms the relative basis as a measure of creativity, and stratified
4
5 sampling offers a means to achieve this diversity of artwork.
6
7

8 9 **2.1 Research questions and hypothesis**

10 The study used a creative typographic task that was developed in Jeffries (2015). This
11
12 Type Task was based on the seminal graphic design book *Watching Words Move*
13 (Chermayeff & Geismar, 2006) and, as a design exercise, has longstanding and current
14
15 usage in design education. The Type Task uses text only, and requires the participant to
16
17 choose a word, and then visually communicate that word through the use of type. For
18
19 instance, in figure 1 the word "Saw" is written by extending the last letter "W" to
20
21 suggest the teeth of a saw, or the word "Imagine" has the middle letter deleted, leaving
22
23 the viewer to imagine what this deletion may be.
24
25
26
27
28



36 Figure 1. Two examples of the artwork related to the Type Task (neither example was
37 part of this study)
38
39
40
41
42
43
44
45
46
47

48 Building on the 2015 study, the current study was guided by two research questions:

- 49 • Can the Type Task remain a reliable measure of graphic design creativity
50 without pre-selection of artwork?
51
52
- 53 • To what extent are technical execution and aesthetic appeal correlated with
54 graphic design creativity?
55
56
57
58
59
60

1
2
3 The expectation (given the finding of Jeffries, 2015) was that Cronbach's alpha for
4 graphic design creativity would remain above 0.7, and, based on previous CAT studies
5 focused on design creativity (such as: Christiaans, 2002; Wojtczuk & Bonnardel, 2012;
6 Valgeirsdottir, Onarheim & Gabrielsen, 2015), that there would be a moderate to strong
7 significant correlations between judges rating of technical execution, aesthetic appeal,
8 and creativity, in most, if not all, instances.
9
10
11
12
13
14
15
16
17

18 **3. Method**

19 CAT studies require two broad groups of participants: those who would generate the
20 creative outputs (participants), and those who would assess the creative outputs
21 (judges). Creative output (in this case, graphic design artwork) was judged by eight
22 professional graphic designers, and a range of statistical analysis methods were used as
23 required, Cronbach's Alpha, Pearson's Correlations Coefficient and Factor Analysis.
24
25
26
27
28
29
30
31
32

33 3.1 Participants

34 While participants were stratified between novices, intermediates and experts, to offer a
35 range of graphic design creativity, it was important they were considered as one group,
36 as a whole, for data analysis. For some researchers, thirty participants are considered to
37 be a threshold for the use of parametric statistics (Gall, Borg & Gall, 1996); for others,
38 this is a topic of debate and would prefer to set a baseline of 50 participants. For this
39 reason, a power analysis was performed and highlighted that with a sample size of 32
40 participants, a one-tailed correlational analysis would have 80% power to detect a
41 moderate correlation if it was present in the data.
42
43
44
45
46
47
48
49
50
51
52
53
54

55 Before undertaking the creativity tasks, each participant's level of experience in graphic
56 design was checked through a short questionnaire after they had received a consent
57
58
59
60

1
2
3 form and information sheet. The study was granted ethical approval on the basis that
4
5 levels of experience would not be correlated with creativity scores. Each level of
6
7 expertise was identified by gaining information as follows:
8
9

- 10
11 1: I work as a full-time graphic designer, and have done so for more than five years
- 12
13 2: I have worked as a designer, artist or craftsman in the last ten years.
- 14
15 3: I am currently studying a full-time graphic design degree, and have done so for more
16
17 than two years
- 18
19 4: I have studied design, art or craft either formally or as a hobby in the last ten years.
- 20
21 5: I have not studied design, art or craft either formally, or as a hobby, since leaving
22
23 secondary school.
24
25
26
27

28
29 After completion of the study, participants had the chance to ask questions and were
30
31 given a £10 Amazon voucher as a gesture of thanks for their support. In total 118,
32
33 professional graphic designers, student graphic designers and members of the general
34
35 public were contacted. 32 took part and completed the Type Task; 18 females and 14
36
37 males, with a mean age of 32.69 years (SD: 11.27), that ranged from 19 to 58 years of
38
39 age.
40
41
42

43
44 Novices: ten females and one male. The mean age was 35.27 (SD: 7.95 years) and
45
46 ranged from 23 to 49. Seven participants had not studied design, art or craft either
47
48 formally, or as a hobby, since leaving secondary school. Four participants said they had
49
50 studied design, art or craft either formally or as a hobby in the last ten years. Example
51
52 comments for this rating were: "I do a lot of crafts as a hobby. Crochet, jewellery", or "I
53
54
55
56
57
58
59
60

1
2
3 have done a little silversmithing and bookbinding as a hobby - more technical than
4
5 design. No formal study, or art or graphics at all."
6
7

8
9 Intermediates: six females, and four males. The mean age was 20.70 years (SD: 1.34)
10
11 and ranged from 19 to 24. All described themselves as currently studying a full-time
12
13 graphic design degree, and have done so for more than two years.
14
15

16
17
18 Experts: two females, and nine males. Experience in graphic design ranged from 4 to 35
19
20 years, with a mean of 19 years (SD: 9.77 years). The mean age was 41 (SD: 10.08
21
22 years) and ranged from 26 to 58 years. The majority of experts described themselves as
23
24 a graphic designer, designer or creative; other titles were an art director, creative
25
26 director, or brand engagement specialist.
27
28

29
30
31 The work assessed was created under non-experimental conditions, as it has been shown
32
33 that CAT inter-rater reliability could remain acceptable even when creative outputs are
34
35 not generated under experimental conditions (Baer, Kaufman & Gentile, 2004;
36
37 Christiaans & Venselaar, 2005). The Type Task was completed by participants in their
38
39 own time, and unsupervised by the researcher.
40
41
42

43 44 **3.2 The Type Task**

45
46 The Type Task (see figure 1.) was the same graphic design creativity task used in
47
48 Jeffries, 2015, with some minor adaptations. Of three groups of participants, novices were
49
50 the most challenging group in respect to undertaking the Type Task. Experts and
51
52 intermediates were likely familiar with the Type Task through their professional and
53
54 educational experience, and would certainly be versed in the computer software that
55
56 aided undertaking the Type Task. This was unlikely to be the case for novices.
57
58
59
60

1
2
3
4
5 The probable mismatch in technical software skills between novices and
6
7 intermediate/experts was considered. For example, if only basic computer software
8
9 were used (Microsoft Word and PowerPoint are software packages available to the vast
10
11 majority of working participants), while this may match the technical skill level of the
12
13 novices, it could undermine the creative freedom of the intermediate/experts.
14
15

16
17
18 Perhaps of more importance is that the limitations of such software may restrict the
19
20 creativity for all participants, as the challenge becomes not about one's creative vision,
21
22 but how to implement this within the constraints of the software. In contrast, to specify
23
24 that designs be created on specialist software (such as Adobe Illustrator or InDesign)
25
26 clearly disadvantages novices: they would likely be unfamiliar with such software
27
28 (which is daunting for most at first) and significantly more of their time would be taken
29
30 with understanding how to use the software rather than the creativity of their design.
31
32

33
34 What was required was to reduce this mismatch in technical software skills; to the
35
36 extent that not knowing how to do something in a computer package did not impinge on
37
38 the creativity of the concept, and its resulting manifestation to the satisfaction of the
39
40 participant. As two-thirds of the participants would be intermediate or experts, the
41
42 decision was taken to use Adobe Illustrator, and not to restrict the software they would
43
44 have available for the Type Task to basic software.
45
46
47

48
49 To make sure that novices were not disadvantaged, the solution was to give novices
50
51 access to software technical support that enabled them to realise their ideas. In this
52
53 respect, novices may not know how to produce their ideas in Adobe Illustrator but
54
55 would be able to dictate their ideas to a technician who could. Exactly how this support
56
57
58
59
60

1
2
3 was provided, for how long, and how to guard against software support slipping into
4
5 design suggestions, and thus not reflecting the creativity of the participant, was
6
7 considered. For example, decisions on layout, type, and choice of work would belong to
8
9 the novice, with the technical role purely to transfer ideas into Adobe Illustrator. Such
10
11 conditions were presented to participants through one A4 sheet on technical support;
12
13 relevant details are highlighted as follows:
14
15

16 *As this Type Task is designed to offer a creative challenge across the full range of*
17 *graphic design expertise, it is important that the tools available to create artwork*
18 *provide the greatest freedom of expression...The technical support is to enable each*
19 *participant, regardless of expertise with Adobe Illustrator, to create an artwork that*
20 *reflects his or her creative vision. It is vital, however, that this technical support only*
21 *deals with the practical aspects of transferring a participants design to Adobe*
22 *Illustrator, rather than offering advice or suggestions on the quality of the design...To*
23 *safeguard against such influence, where possible, technical support will be offered*
24 *remotely via email. For example, a participant may send a hand drawn illustrations of*
25 *their design via email to kjeffries@uclan.ac.uk. This hand drawn illustration will then*
26 *be transferred to Adobe Illustrator, and this digital illustration will be sent back to the*
27 *participant. Any amendments required by a participant can be suggested, and the cycle*
28 *will continue until the participant is satisfied, or until the time limit on the task has*
29 *expired...The time required creating this artwork is expected to take no more than one*
30 *hour. However, thinking time and playing with ideas may increase for some*
31 *participants...This task is not timed in a conventional sense but for practical purposes,*
32 *a limit of one week from receiving the task has been set...Unless otherwise stated by a*
33 *participant, the default font will be Gill Sans, and the size of the type will be selected to*
34 *fit an A4 sheet (allowing for a 20 mm border).*
35
36
37
38

39 Before beginning the recruitment of participants for this study, the Type Task was
40
41 piloted with three novices. Importantly, the application of technical support did not
42
43 appear to influence the design a participant had visualised. Through photographs of
44
45 sketches, and written instructions novice participants were more than capable of
46
47 expressing their creative vision, and ask for amendments until the design was to their
48
49 satisfaction.
50
51
52
53
54
55
56
57
58
59
60

3.3 Judges and Rating Procedures.

Eight judges independently rated the artwork using the CAT procedures outlined in Jeffries (2015) and discussed further in this section. From these eight judges, four were females, and four were males. The age range was between 31 and 55; the mean was 41.63 years (SD of 8.88 years). Years of experience in professional graphic design ranged from 9 to 34 years; the mean was 19.13 years, (SD of 8.43 years).

After signing the consent form, each judge rated the 32 artworks for creativity only (i.e. discounting technical execution or aesthetic appeal from their creativity rating).

Specifically, the creativity only instructions were worded as follows:

Instructions for Judges: How to rate these artworks

There is only one criterion in rating these artworks: creativity. We realise that creativity probably overlaps other criteria one might consider (for example: aesthetic appeal, or technical execution) but we ask you to rate the artworks solely on the basis of their creativity. There is no need to explain or defend your ratings in any way; we ask only that you use your own sense of which is more or less creative (**relative to the other artworks provided**).

Please look through these artworks three times, and rate them for creativity.

- The first time familiarise yourself with all the artworks provided.
- The second time, group the artworks into Low, Medium, or High ratings.
- The third time, assign a numerical rating between 1 and 6 (1's being the least creative and 6's being the most creative).

There should be a roughly even number of artworks at each of the six levels. It is very important that you use the full 1-6 scale.

The design brief was given to the judges, and they were told this was the same brief seen by the participants, and that at the back was the instruction on technical support that was provided to all participants. Judges were then left to read these documents and the researcher set up the laptop to view the artwork. The order of artwork was

randomised, and they were free to control how long they viewed artworks and could return to each artwork for further inspection. Each judge familiarised themselves with all the artworks, and when satisfied informed the researcher, they were ready to continue. Judges were given an A3 laminated rating sheet, see Figure 2 (developed to graphically reinforce the CAT protocol and instructions), and a set of laminated cards. These cards were miniature copies of the artwork they had just viewed. Cards were placed in a stack, by the researcher, onto the rating sheet area designated “Medium”, and judges proceeded to rate the artwork, and had as much time as they required.

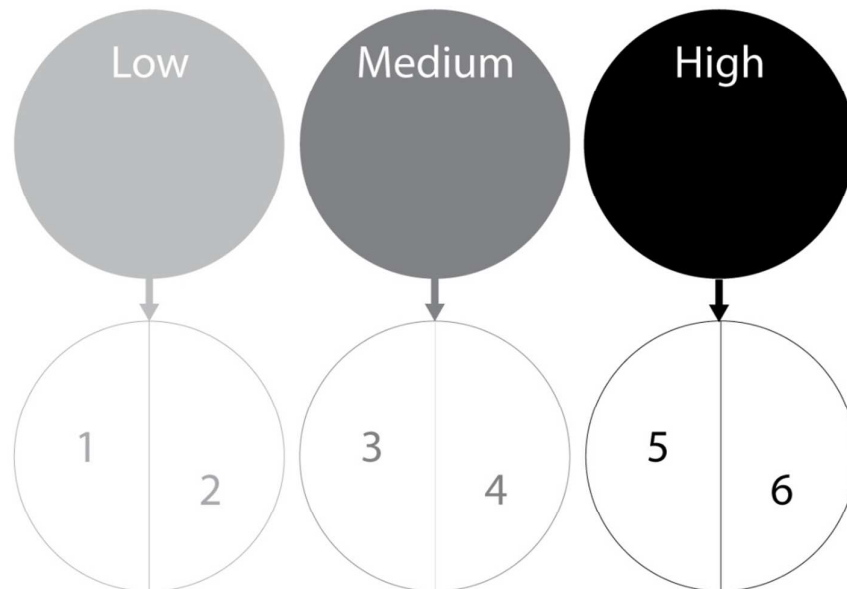


Figure 2. Rating sheet developed to graphically reinforce the CAT protocol

Once judges had read the instructions, three points were re-emphasized within the instructions. First, judges were to rate the artwork on creativity only (without considering technical execution and aesthetic appeal). This instruction is especially important to emphasise and draws a distinction between the CAT protocol in Jeffries (2015) with the present study. By verbally re-emphasizing this aspect of the instructions it could be argued that this additional procedure addresses whether a judge had read the

1
2
3 instructions sufficiently and whether the wording alone was robust enough to get the
4
5 point across. Second, ratings were relative to the artworks provided, rather than an
6
7 external standard of creativity (such as award-winning creativity). Third, it was
8
9 important that the judges should divide the artworks more or less evenly over the ratings
10
11 1 to 6. The even number was to emphasise that judges use the whole of the 1 to 6 rating
12
13 scale, and not overuse some of the scale (i.e. award everything a 1). A skew in rating
14
15 can impact on the reliability of Cronbach's alpha, and would not be in keeping with the
16
17 theoretical premise of the CAT being a relative measure of creativity within a defined
18
19 sample.
20
21
22
23
24

25 Once the creativity assessment was completed, judges were given the instruction to rate
26
27 either technical execution or aesthetic appeal. The previous creativity ratings were
28
29 removed, the artworks were randomly re-sorted, then placed back to the top of the
30
31 rating sheet. Judges were left to begin their assessment. With this completed, the same
32
33 procedure was followed but for the remaining instructions, either aesthetic appeal or
34
35 technical execution, which was determined by the randomization of judges into
36
37 counterbalanced conditions. The wording of instruction for rating aesthetic appeal (see
38
39 below) was virtually identical to that of technical executions: the only change was to
40
41 replace the phrase aesthetic appeal with that of technical execution.
42
43
44
45
46
47

48 Aesthetic Appeal

49 Please look through these artworks and rate them for aesthetic appeal. There is no need
50
51 to explain or defend your ratings in any way; we ask only that you use your own sense
52
53 of which is more or less aesthetically appealing (**relative to the other artworks
54
55 provided**).

- 56 • Familiarise yourself with all the artworks provided.
- 57 • Group the artworks into Low, Medium, or High ratings.
- 58
- 59
- 60

- Assign a numerical rating between 1 and 6 (1's being the least aesthetically appealing and 6's being the most aesthetically appealing).

There should be a roughly even number of artworks at each of the six levels. It is very important that you use the full 1-6 scale.

4. Results

As discussed above, the expectation was that the inter-rater agreement for creativity would give a Cronbach's alpha above 0.7 and that there would be a moderate to strong significant correlation between technical execution, aesthetic appeal, and graphic design creativity.

The Inter-rater agreement for Creativity was a Cronbach's alpha of 0.92; for Aesthetic appeal, 0.88; Technical execution, 0.88. Each rating was above the 0.7 alpha level commonly used as a threshold for an appropriate standard of consensus among judges. The combined scores for each of the three ratings was computed and to check for normality of distribution; histograms were plotted for each rating. Skewness and standard error of skew were calculated to check if data would satisfy the assumptions of correlation coefficient analysis. All Z-skew statistics were within the threshold of +/- 1.96, suggesting the data was appropriate for parametric analysis. On this basis, Pearson's r was used to calculate the correlation coefficient between each rating. All correlations were positive at a significance level of 0.01 (1 tailed), and each suggestive of very strong correlations (Figure 3): creativity and aesthetic appeal (r: 0.93); aesthetic appeal and technical execution (r: 0.90); creativity and technical execution (r: 0.87).

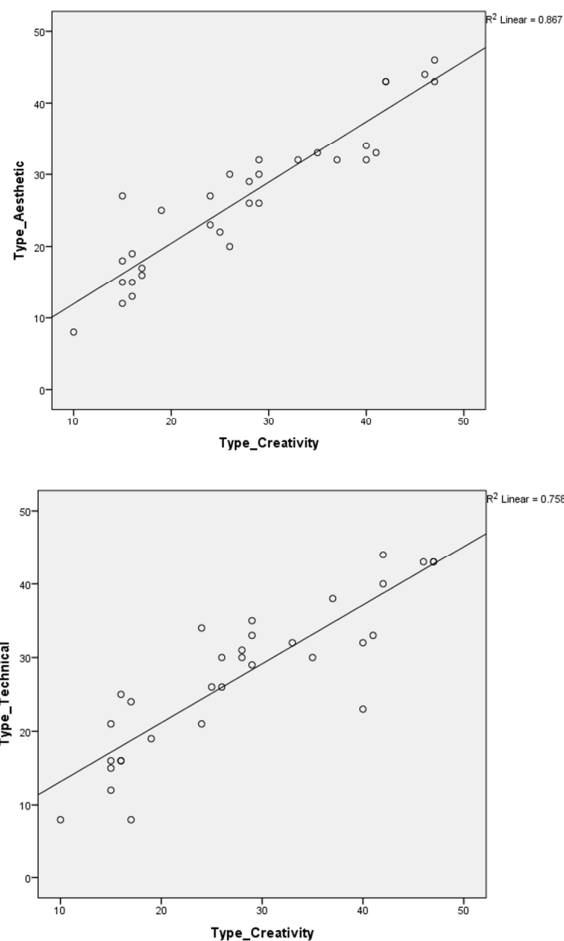


Figure 3. Scatter plots and correlations of creativity, technical execution, and aesthetic appeal.

In addition to the correlations above, a Principal Components Analysis (PCA) on the correlation coefficients was computed. The total variance, Scree plot, and component matrix are shown in Table 1 and 2 and Figure 4. Evidence can be found that a substantial single component is likely to be present which underpins judges' assessment of creativity, technical execution, and aesthetic appeal.

For the Type Task, only one principal component with an eigenvalue greater than one was identified, and the scree plot is characteristic of a data set with a single principal component. This single component explains 93% of the variance in the data.

Table 1: PCA of Type Task

Total Variance Explained						
Component	Initial Eigenvalues			Extraction Sums of Squared Loadings		
	Total	% of Variance	Cumulative %	Total	% of Variance	Cumulative %
1	2.803	93.430	93.430	2.803	93.430	93.430
2	.133	4.440	97.870	.133	4.440	97.870
3	.064	2.130	100.000	.064	2.130	100.000

Extraction Method: Principal Component Analysis.

Table 2: Type Task PCA Component Matrix

Component Matrix ^a			
	Component		
	1	2	3
CREATIVITY	.966	-.212	.145
AESTHETIC	.977	-.073	-.199
TECHNICAL	.956	.288	.057

Extraction Method: Principal Component Analysis.

a. 3 components extracted.

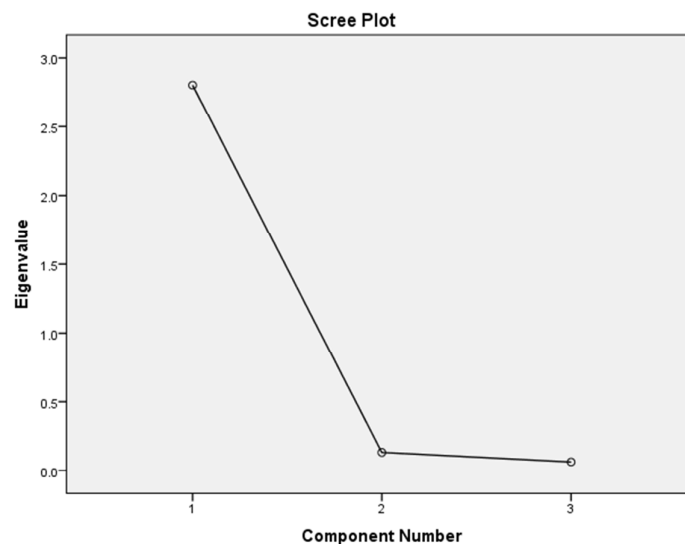


Figure 4: Scree plot for Type Task

5. Discussion

As highlighted in Jeffries, 2012, despite extensive use within creativity research, the use of the CAT as a measure of creativity within design research was relatively small.

1
2
3 Indeed, the study evidenced that only two design journal papers had operationalized
4
5 CAT during a thirty-year period. Jeffries (2015) identified that the CAT could be a
6
7 reliable measure of graphic design creativity, but consideration of a number of caveats
8
9 was prudent, and the findings here suggest this has remained the case. Inter-rater
10
11 agreement for creativity (α : 0.92) was considerably more than the 0.7 threshold, and
12
13 a higher level of consensus than in Jeffries, 2015, (0.73), which ran the same tasks and
14
15 had the same number of judges. This comparison, however, needs to be considered
16
17 alongside the fact that judges were different, as was the artwork being judged. A further
18
19 point was the population of participants in this study was more diverse than the 2015
20
21 study, and this may have impacted on the higher level of consensus.
22
23
24
25
26

27 Notwithstanding the limitations of a direct comparison between Jeffries (2015), and
28
29 here, the increase in judges' consensus on creativity from 0.73 (in the 2015 study) to
30
31 0.92 (in this study) is likely to be significant. With more research using the Type Task,
32
33 future experimental studies on CAT protocol will be able to identify the exact extent
34
35 upon which this increase in alpha level is likely to occur. On the basis of published
36
37 graphic design creativity studies, 0.92 is the highest known alpha to date.
38
39
40
41
42

43 Adjustments of CAT protocol were made in this study that may explain this increase
44
45 beyond the influence of sampling bias/error. Notably, the verbal re-emphasis to judges
46
47 not to compound their rating of creativity with technical execution or aesthetic appeal.
48
49 If this is the case, then it lends further support towards the benefit of "creativity only"
50
51 instructions, the value of verbal re-emphasis, and diversity of artwork. These caveats to
52
53 CAT protocol do not seem to be detrimental to research design and would appear to
54
55 improve the reliability, if not the validity, of CAT as a measure of creativity.
56
57
58
59
60

1
2
3
4
5 It is for this reason that the level of attention to procedure and precision of details
6
7 regarding the CAT's application to graphic design creativity are present. For some, this
8
9 degree of detail may be laboured (especially given such discussion is not always present
10
11 in other scholars application of the technique), but the argument is that such details do
12
13 matter. It is possible that creativity assessment in graphic design is particularly sensitive
14
15 to such nuances of the method, and it may be that in other domains this degree of detail
16
17 is not required. However, as is the case in other areas of design (like ergonomics) the
18
19 value of designing for limiting users offers the opportunity to create a better "product"
20
21 for all: the 95 percentile who fit within a population, as well as those found within the
22
23 extremes.
24
25
26
27

28
29 As detailed in the method section, judges were asked first to assess creativity only, and
30
31 then randomly assigned to judge technical execution or aesthetic appeal. A final
32
33 assessment was either technical execution or aesthetic appeal depending on the previous
34
35 outcome.
36
37
38
39

40
41 For some time, guidance has been that when a CAT task is developed for a different
42
43 domain (i.e. one that has not been studied with a particular task), researchers should ask
44
45 judges to rate both technical execution and aesthetic appeal, and check to see creativity
46
47 ratings are distinct from these criteria. For the Type Task, each correlation was positive,
48
49 significant and each suggestive of very strong correlations. It would appear that the
50
51 Type Task is an obvious example of where creativity, technical execution, and aesthetic
52
53 appeal are highly correlated.
54
55
56
57
58
59
60

1
2
3 In each of the scatter plots shown in the Results section (see Figure 3), and through the
4
5 PCA, evidence can be found that a single substantial component is likely to be present
6
7 that underpins the judges' assessment of creativity, technical execution, and aesthetic
8
9 appeal. If this is the case for other design-focused CAT studies, then two key questions
10
11 arise.
12

13
14
15
16 Firstly, is this original guideline on the distinction between creativity, technical
17
18 execution and aesthetic appeal correct? Again, it must be acknowledged that Amabile
19
20 and her colleagues made very clear this would not apply to some domains. Equally, the
21
22 CAT was foremost used in social psychological research contexts, and showing
23
24 discriminant validity between creativity and other compounding constructs was
25
26 important to establish the validity of the measure at that time. Never the less,
27
28 contradictory evidence in Amabile and Hennessey's work raises doubts about creativity
29
30 and technical execution as distinct constructs. Within Amabile's 1982 studies she found
31
32 correlations as high as .77 between creativity and technical goodness, and Hennessey
33
34 (1994) presented statistically significant correlations as high as .71.
35
36
37
38
39

40
41 As identified in the Background section of this paper, other contemporary CAT research
42
43 studies directly related to graphic design have identified high positive correlations
44
45 (Wojtczuk & Bonnardel, 2012) between creativity and aesthetics. Equally, CAT
46
47 inspired design studies exploring the creativity of product designs (Valgeirsdottir,
48
49 Onarheim & Gabrielsen, 2015) have evidenced strong positive correlations between
50
51 creativity and aesthetic appeal. For Industrial Design Engineering students, Christiaans'
52
53 (2002) found strong correlations between creativity and aesthetic criteria, but not for
54
55 creativity and technical quality.
56
57
58
59
60

1
2
3
4
5 In contrast, some design creativity researchers, when introducing a new CAT design
6 task gather ratings for creativity, technical execution and aesthetic appeal but, given the
7 focus of their study, do not report the correlations between these (Yuan & Lee, 2014);
8 some design researchers using new CAT design tasks appear not to have gathered
9 ratings from judges on technical execution or aesthetic appeal alongside creativity
10 assessments. For example, it is unclear if Meneely and Portillo's 2005 study (with
11 interior design students who undertook a furniture design task) had rated technical
12 execution and aesthetic appeal in previous studies using this task. Given the range of
13 interpretations of CAT protocols for design tasks, it would suggest the original guidance
14 on evidencing a distinction between creativity, technical execution and aesthetic appeal
15 is problematic for design creativity research.
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31

32 Secondly, what does it mean if creativity, technical execution and aesthetic appeal are
33 highly correlated? Moreover, what happens in those design domains where they are
34 correlated: should researchers not use CAT? On one level it may be possible for
35 detractors of CAT to see this evidence of confluence as a basis to say the method does
36 not measure creativity. This is not the position of this paper: the argument here is that
37 assessing creativity output is complex (especially graphic design creativity, and likely
38 for other areas of design). Indeed, complexity is the reason why CAT did not work
39 when first applied to graphic design. With various caveats considered, clearly,
40 acceptable levels of consensus can be achieved. One of those caveats is how important
41 it is to instruct judges not to consider technical execution and aesthetic appeal in their
42 ratings of creativity. The case made here is further support for why the caveat on
43 technical executions and aesthetic appeal is important. These constructs would appear to
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

1
2
3 be highly correlated for certain domains (graphic design for one), but this does not mean
4
5 that judges cannot give a rating on creativity only; what is required is that this
6
7 instruction is made clear to them. Emphasising that caveat enables a judge to tacitly
8
9 reduce the background "noise" in creativity rating that is present from technical
10
11 execution and aesthetic appeal. Isolating creativity may be a challenge, but the evidence
12
13 is that they can achieve this. In contrast, if researchers do not do this, then every study
14
15 could be skewing the creativity aspect with technical execution and aesthetic appeal.
16
17 That has implications for comparison across studies.
18
19

20
21
22
23 It is worth highlighting that the CAT is considered to be a theoretically neutral measure
24
25 of creativity, in that suitable judges are free to use whatever tacit or explicit criteria they
26
27 may have evolved regarding creativity within their domain. In this respect, the CAT is
28
29 quite distinct from other measures of creativity that define the criteria judges should
30
31 apply when rating creativity: for example novelty, surprise, or usefulness. This
32
33 theoretical neutrality is one reason creativity researchers have advocated the CAT as a
34
35 "Gold Standard" of creativity assessment (Kaufman, Plucker & Baer, 2008).
36
37
38
39

40
41 The focus of this research study, and the previous study in 2015 is the influence of CAT
42
43 instructions and protocols on the reliability of assessing graphic design creativity. The
44
45 argument presented is that in domains (like graphic design) where the relationship
46
47 between creativity, aesthetic appeal and technical execution are highly correlated, it is
48
49 valuable for design researchers applying the CAT to use creativity only instructions and
50
51 verbally re-emphasise these to maintain its theoretical neutrality. Without a researcher
52
53 being explicit on creativity only ratings, the preferences of some judges for aesthetic
54
55 appeal or technical executions would appear to impact on the reliability of their
56
57
58
59
60

1
2
3 creativity assessments. As an international design research community, our further
4 exploration of the science of creativity now requires us to harmonise the tasks,
5 instructions and protocols we used for the CAT. This study and the 2015 study are
6 ultimately about what those CAT instructions and protocols may be for design creativity
7 researchers.
8
9
10
11
12

13 14 15 16 **5.1 Emergent considerations**

17 One issue that emerged from the analysis was the level of detail for a graphic design
18 task/brief. For example, the Type Task did not clarify which format the A4 sheet should
19 be presented in; the majority of artworks from participants were landscape, while a few
20 others chose portrait. To what extent such variation in format, or in some cases the
21 layout of the graphic on the sheet, may influence the judges' consensus on creativity
22 appears not to have had a direct impact overall on consensus, but it is likely that such
23 choices could have had an influence on views of technical execution and/or aesthetics
24 appeal.
25
26
27
28
29
30
31
32
33
34
35
36
37

38 From a design perspective, such choices by participants to opt for landscape or portrait
39 was not problematic. The challenge came from the administration of the CAT. In the
40 creation of a digital file to show judges, the artwork in portrait format was compounded
41 by the default setting and limitations of PowerPoint. These constraints had not been
42 apparent in the previous studies. Unfortunately, it was not straightforward to switch
43 between landscape and portrait in a single PowerPoint file. For this reason, Adobe
44 Illustrator was used instead, however, having created a PDF with both landscape and
45 portrait artwork shown correctly, the challenge of randomising the artwork in Adobe
46 Illustrator was more difficult. Artboards for each artwork in Adobe Illustrator needed to
47
48
49
50
51
52
53
54
55
56
57
58
59
60

1
2
3 be reordered by hand for each judge. This process appears not to be amenable to
4
5 automation, and while laborious did achieve the PDF file showing a random order of
6
7 artwork, and in the appropriate page orientation. Whether such additional procedures
8
9 are worth the effort of offering participants the opportunity to select portrait and
10
11 landscape formats is something to consider for future studies and discussion. For
12
13 graphic design creativity the restriction to, for example, landscape only formats, for
14
15 future research may be considered too restrictive a condition within the art and design
16
17 community.
18
19

20
21
22
23 A final point on the issue of portrait or landscape format came from observing the
24
25 judges as they undertook their assessments. Judges had four elements to engage with: an
26
27 A4 hard copies of the design brief and instruction for rating the work; a laptop showing
28
29 a PDF file that had all the artwork shown in the correct format (either landscape or
30
31 portrait); an A3 laminated rating sheet; a set of laminated thumbnails of the artworks
32
33 shown on the laptop that they could move around on the A3 rating sheet. What was
34
35 noticeable, was that some artworks shown as portrait would occasionally find
36
37 themselves in a landscape position on the rating sheet. Equally, as the rating progressed
38
39 from creativity to technical execution, to aesthetic appeal, for example, judges tended to
40
41 spend less time looking at the PDF file on the laptop and more time concentrating on
42
43 the rating sheet. It was for this reason that each time after a judge had first assessed the
44
45 creativity of the artworks, they were reminded that the details, both technical and
46
47 aesthetic, would be better found on screen than in the thumbnails. It is perhaps a minor
48
49 point given the high levels of consensus achieved in this study, but some artworks that
50
51 were portrait may have suffered from being viewed as landscape thumbnails and judged
52
53 more harshly. Especially, in those instances where the portrait format was crucial to
54
55
56
57
58
59
60

1
2
3 understanding the work, it is difficult to rule out they may not have been disadvantaged
4
5 by the paper based rating sheet, and the dominance of predominantly portrait format
6
7 artworks. Whether this is enough of a rationale to standardise the Type Task to a
8
9 landscape only format, is difficult to say.
10

11
12
13
14 This issue may appear overly detailed, but it reflects a broader point regarding how
15
16 standardized should a CAT graphic design creativity task be, or could become: did the
17
18 current instructions give an opportunity for creativity because they did not specify a
19
20 standard format to be followed by all participants, or does the lack of a standard format
21
22 introduce more complexity and distract judges unnecessarily? It was possible to
23
24 standardise all the artworks judges would view through opting for a landscape format.
25
26 However, while this may enable easier administration and allow judges to view more
27
28 easily all the artwork without the distraction of changes in format, for some artworks the
29
30 choice of landscape or portrait was part of the quality of the work. The point highlights
31
32 that while such consideration may not be of great importance in some domains, a
33
34 graphic design CAT appears to offer a unique testing ground for design creativity
35
36 researchers, one that is highly sensitive and responsive to the refinements of our
37
38 methods.
39
40
41
42
43
44
45

46 **6. Conclusion**

47
48 Given the findings of this paper, it is worthwhile to reflect on the broader value of
49
50 measuring technical execution and aesthetic appeal, alongside creativity. If assumptions
51
52 of distinction are not met, but we have instructions and protocols that minimise the
53
54 background "noise" of technical execution and aesthetic appeal on creativity
55
56 assessment, then is rating additional constructs really worth the extra time and effort of
57
58
59
60

1
2
3 future design researchers who use the CAT? For example, Valgeirsdottir, Onarheim &
4
5 Gabrielsen (2015) asked judges to rate three other constructs and creativity; the study
6
7 above asked for two other constructs. In each study, rating additional components is a
8
9 considerable amount of work when the reason for using the CAT is to assess creative
10
11 output.
12

13
14
15
16 These are issues to be debated within the broader community of scholars. For the value
17
18 of starting that debate, the conclusion of this paper is that if design creativity researchers
19
20 want only to measure creativity, they should do so, and without this being viewed as a
21
22 methodological flaw because of the original guidelines for the CAT. Obviously, such a
23
24 statement is dependent on the research aims of a given study, but the arguments are that
25
26 investment in a judge's time and effort, and the researcher's time and effort, make
27
28 additional assessments unproductive. Recruitment to CAT studies is an acknowledged
29
30 challenge as it currently stands; when the likelihood is there will be a highly significant
31
32 and strong correlation between creativity, technical execution, and aesthetic appeal, it is
33
34 questionable if that additional effort is worthwhile for the majority of future design
35
36 creativity studies
37
38
39
40
41

42 **Acknowledgments**

43 removed
44

45 **References**

- 46
47 Amabile, T. M. (1982). Social psychology of creativity: A consensual assessment technique. *Journal of*
48 *personality and social psychology*, 43, 997-1013.
49
50 Hennessey, B. A., Amabile, T. M., & Mueller, J. S. (2011). Consensual assessment. In M. A. Runco & S.
51 R. Pritzker (Eds.), *Encyclopedia of creativity* (2nd Ed., Vol. 1) (pp. 253-260). San Diego, CA:
52 Academic Press.
53
54 Baer, J., Kaufman, J. C., & Gentile, C. A. (2004). Extension of the Consensual Assessment Technique to
55 Nonparallel Creative Products. *Creativity Research Journal*, 16, 113-117.
56
57
58
59
60

- 1
2
3 Baer, J., & McKool, S. S. (2009). Assessing creativity using the Consensual Assessment Technique. In
4 *Handbook of research on assessment technologies, methods, and applications in higher*
5 *education* (pp. 1–13).
6
- 7 Chermayeff, I., & Geismar, T. H. (2006). *Watching Words Move*. Chronicle Books.
- 8 Christiaans, H. (2002). Creativity as a Design Criterion. *Creativity Research Journal*, 14, 41-54
- 9 Christiaans, H. & Venselaar, K. (2005). Creativity in design engineering and the role of knowledge:
10 Modelling the expert. *International Journal of Technology and Design Education*, 15, 217-236.
- 11 Cseh, G. M. (2014). *Flow in visual creativity* (Unpublished doctoral dissertation). University of
12 Aberdeen, Scotland.
- 13 Gall, M. D., Borg, W. R., & Gall, J. P. (1996). *Educational research: An introduction*. White Plains, NY:
14 Longman.
- 15 Hennessey, B. A. (1994). The consensual assessment technique: An examination of the relationship
16 between ratings of product and process creativity. *Creativity Research Journal*, 7, 193-208.
- 17 Jeffries, K. K. (2012). Amabile's Consensual Assessment Technique: Why has it not been used more in
18 design creativity research? *Proceedings of the 2nd International Conference on Design*
19 *Creativity (ICDC2012)*, 1, 211-220
- 20 Jeffries, K. K. (2015). A CAT with caveats: Is the Consensual Assessment Technique a reliable measure
21 of graphic design creativity? *The International Journal of Design Creativity and Innovation*.
- 22 Kaufman, J. C., Plucker, J. A., & Baer, J. (2008). *Essentials of creativity assessment*. Hoboken, NJ US:
23 John Wiley & Sons Inc.
- 24 Meneely, J., & Portillo, M. (2005). The adaptable mind in design: Relating personality, cognitive style,
25 and creative performance. *Creativity Research Journal*, 17(2-3), 155-166.
- 26 Valgeirsdottir, D., Onarheim, B., & Gabrielsen, G. (2015). Product creativity assessment of innovations:
27 considering the creative process. *International Journal of Design Creativity and Innovation*, 3,
28 95-106.
- 29 Wojtczuk, A., & Bonnardel, N. (2012). Differences in creative design assessment. *Proceedings of the 2nd*
30 *International Conference on Design Creativity (ICDC2012)*.
- 31 Yuan, X., & Lee, J. H. (2014). A quantitative approach for assessment of creativity in product
32 design. *Advanced Engineering Informatics*, 28(4), 528-541.
- 33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60