This article was downloaded by: *[Massachusetts Institute of Technology]*
On: *18 October 2009*
Access details: *Access Details: [subscription number 906464818]*
Publisher *Routledge*
Informa Ltd Registered in England and Wales Registered Number: 1072954 Registered office: Mortimer House, 37-41 Mortimer Street, London W1T 3JH, UK

Creativity Research Journal
Publication details, including instructions for authors and subscription information:
http://www.informaworld.com/smpp/title~content=t775653635

Creativity as a Design Criterion

Henri H. C. M. Christiaans [a]
[a] Delft University, The Netherlands.

Online Publication Date: 01 January 2002

To cite this Article Christiaans, Henri H. C. M.(2002)'Creativity as a Design Criterion',Creativity Research Journal,14:1,41 — 54
To link to this Article: DOI: 10.1207/S15326934CRJ1401_4
URL: http://dx.doi.org/10.1207/S15326934CRJ1401_4

PLEASE SCROLL DOWN FOR ARTICLE

Full terms and conditions of use: http://www.informaworld.com/terms-and-conditions-of-access.pdf

This article may be used for research, teaching and private study purposes. Any substantial or systematic reproduction, re-distribution, re-selling, loan or sub-licensing, systematic supply or distribution in any form to anyone is expressly forbidden.

The publisher does not give any warranty express or implied or make any representation that the contents will be complete or accurate or up to date. The accuracy of any instructions, formulae and drug doses should be independently verified with primary sources. The publisher shall not be liable for any loss, actions, claims, proceedings, demand or costs or damages whatsoever or howsoever caused arising directly or indirectly in connection with or arising out of the use of this material.

Creativity Research Journal
2002, Vol. 14, No. 1, 41–54

Copyright 2002 by
Lawrence Erlbaum Associates, Inc.

Downloaded By: [Massachusetts Institute of Technology] At: 07:11 18 October 2009

Creativity as a Design Criterion

Henri H. C. M. Christiaans
Delft University, The Netherlands

ABSTRACT: One of the most important criteria for performance quality in both art and design seems to be the creativity of the product. Being original and innovative is by definition a feature of both areas. The primary objective of this study was to determine whether human judgment of creativity is a reliable and valid method in design evaluation and selection. In a first experiment, the judgments of experts, nonexperts, and people with an intermediate level of expertise were compared. They rated 44 first-year designs on creativity, prototypical value, attractiveness, interest, technical quality, expressiveness, and integrating capacity. Pearson product–moment correlations for creativity were relatively low, ranging from .23 to .29. There was little difference between experts and nonexperts. The results confirmed the research in artwork assessment. In Experiment 2 the results were replicated with senior design students as judges, a group with an intermediate level of expertise. Ratings were given for 3 different designs. Correlations were much higher, ranging from .48 to .57. This could be a consequence of the homogeneity of the group of judges. The prototypicality of a design, the distance between the design and the observers' internal representation, appeared to discriminate between creativity and other aesthetic criteria. A pair-comparisons analysis also contributed to the definition of creativity in both general and domain-specific terms.

In discussions about the quality of a design and of a designer, the concept of creativity is a dominant factor. In addition to fulfilling the functionality criterion, the result of a design activity is often expected to be original, adding value to the existing world of design. In the selection of designs for production in companies, for design awards, and in the field of design education, creativity assessment relies on human judgments. This article raises the question of whether creativity in product design can be judged in a valid and reliable way. The enormous amount of research done in the last decennia highlights the lack of objective methods of assessment. One reason for this lack is that the need for objectivity by formalizing the measurement leads to a reduction of the product features that are appraised (Hofstee, 1985). Features that are difficult to determine will, on the whole, be neglected. Another possible reason is that, because it is made up of aspects such as originality and unexpectedness, the core of the creativity concept cannot be formalized into an objective instrument. Such concepts as creativity and quality have, according to Hofstee, an emergent character; that is, they are defined again and again on the basis of new creations, so that there is no possibility for previous programming. Only a human judge can make ad hoc estimates of the originality of a product. The fact that the human judge is fallible is not a sufficient reason to kick him or her out of the system. When estimating the creativity of products we have to rely on human judgment. In all studies thus far the question has been how to overcome subjectivity within these assessments.

Reliability and Validity

Most creativity assessment studies, relying on human judgment, have been performed in the domain of art, and only a few in design (Amabile, 1983; Ward & Cox, 1974). The results of the art studies show considerable variation in interrater reliability based on correlations between judges. Because they are at different levels of subjectivity, the question is whether artwork judgment

Correspondence and requests for reprints should be sent to Henri H. C. M. Christiaans, Jaffalaan 9, TU Delft, 2628 BX Delft, The Netherlands.

Downloaded By: [Massachusetts Institute of Technology] At: 07:11 18 October 2009

can be compared with design work judgment. The design of products always builds on previous designs and on the archetype of the designed device. Ensuring that the functionality of the product is recognized by the user often takes precedence over aesthetic values. Therefore, objective judgments would seem to be more possible in design work than in artwork. However, although the judging of designs is daily practice in real life, playing an important role in decisions about production and in the awarding of prizes, no controlled experiments have been found to confirm this assumption. The reliability of intersubjective measurement seems also to depend on the expertise of the judges. In the field of art, professionals or trained observers are presumed to be more reliable than naive observers (Hekkert & van Wieringen, 1996; Runco, McCarthy, & Svenson, 1993). Amabile (1982) argued that "appropriate" (familiar with the domain) observers are able to judge creativity. This would apply to any domain in which creativity is a valuable criterion. The assumption is that, based on general cultural values within a society, consistencies will underlie the assessments of judges (Child, 1970). In the assessment of both the aesthetic preference (Temme, 1983) and the level of creativity of artworks and designs, a higher level of agreement will be shown among people who have similar learning experiences in the area of art or design. Problems have arisen, however, regarding the idiosyncratic standards of professional judges. A number of studies report that in the judging of artworks the level of agreement among lay judges is often higher than among experts (Getzels & Csikszentmihalyi, 1976; Gordon, 1956; Hekkert & van Wieringen, 1996). Runco et al. (1993) suggested that expert judges rely on high-level, esoteric, idiosyncratic standards. This makes for less awareness of differences among artworks than is found in groups of judges with lower expertise. Getzels and Csikszentmihalyi (1976) argued that experts have more difficulty assessing products in terms of their fundamental attributes than judges with an intermediate level of expertise. They assumed that experts are much too involved in objects as aesthetic wholes and therefore consider differentiation between attributes as spurious abstractions (cf. p. 111). Their findings were confirmed by Hekkert and van Wieringen (1996). Correlations between mean ratings on originality and other criteria are much higher among experts than among nonexperts.

The validity of subjective judgment is also open to question. An indication of validity might be that judges apparently have no difficulty in distinguishing between various assessment criteria; however, the results of correlational analysis in several studies do not confirm any clear distinction between them. The aesthetic value of the product seems to be strongly related to originality and creativity (Amabile, 1983; Getzels & Csikszentmihalyi, 1976). Findings regarding the relationship between creativity and technical quality are contradictory. In the study of Trowbridge and Charles (1966), the hypothesis that creativity and technical competence can be separated into two distinct variables is confirmed with a correlation of nearly zero. In contrast, in the studies of Getzels and Csikszentmihalyi (1976) and in most of Amabile's (1983) studies, the relationship between the two is quite strong.

Although many studies show that creativity is interrelated with such concepts as aesthetic appeal, appropriateness, and (technical) quality, some authors still claim that creativity can be considered a separate construct (Amabile, 1982, 1983). In this study we tried to find evidence to prove this assumption—that is, that creativity and other aesthetic criteria are different constructs—by introducing a discriminating variable called *prototypical value.* Based on information processing theory, Purcell (1984) argued that the attractiveness of a stimulus increases the more it resembles the prototypical representation of that stimulus. If this theory holds good, then objects that, because they are original and unexpected, are by definition far from being prototypical representations, will be less attractive because of their divergence from the prototype. Because creativity is also characterized by concepts such as originality, the distance between a creative object and the prototypical representation, based on membership of the category of similar objects, is also by definition large, larger than the distance between the aesthetic appeal and the prototypical value.

Experiment 1

Method

Students and judges. Design works were selected from the 1985 first-year student population of the School of Industrial Design Engineering (IDE). The cardboard scale models of 19 computer cabinets and 25 telephone booths were made as a part of the design course. Each model was photographed in two differ-

ent positions and recorded on slides. These slides were used in the experiment.

Teachers and students with varying design expertise acted as judges: 10 design teachers of IDE, 12 senior students of IDE, and 12 students from the mathematics department from the same university. Only male participants were selected. The design teachers, all professional designers, were not involved in the first-year design course. The judges from both student populations were recruited via announcements on notice boards in both departments. As far as the math group was concerned, only first- and second-year students were recruited, whereas the IDE student judges were doing master's degrees. The student judges received a nominal payment.

Procedure. Each judge was asked to rate all 44 designs individually according to seven criteria:

- Creativity: Judges were asked to use their own definition (cf. Amabile, 1983).
- Technical quality: The extent to which a product meets the necessary technical requirements.
- Attractiveness: Preference for the outward form.
- Interest: The extent to which the product arouses interest or fascinates.
- Expressiveness: The extent to which the product expresses its meaning.
- Integrating capacity: The extent to which the product integrates the underlying aspects of form, function, and construction.
- Goodness of example (this term being used by Purcell, 1984): The extent to which the product is prototypical for its class of products. To exclude the influence of other criteria assessed, it was preferable always to begin with the prototypicality criterion.

Results

Agreement. The level of agreement within groups of judges is defined by calculating mean Pearson product–moment correlations. Table 1 (computer cabinets) and Table 2 (telephone booths) show these within-group correlations on each criterion for the

Table 1. *Mean Pearson Correlations Within Groups and Between Groups Judging 19 Computer Cabinets*

	Within Group			Between Group		
	r_{Teacher} (n = 10)	$r_{\text{IDE Student}}$ (n = 12)	$r_{\text{Math Student}}$ (n = 12)	$r_{\text{Teacher/IDE Student}}$	$r_{\text{Teacher/ Math Student}}$	$r_{\text{IDE- Student/ Math Student}}$
Creativity	.28	.26	.23	.29	.28	.26
Prototypical Value	.21	.21	.19	.18	.18	.23
Attractiveness	.17	.12	.18	.15	.17	.07
Interest	.18	.21	.15	.19	.16	.11
Technical Quality	.24	.22	.31	.21	.19	.24
Expressiveness	.15	.08	.21	.12	.05	.08
Integrative Capacity	.20	.04	.24	.12	.09	.09

Table 2. *Mean Pearson Correlations Within Groups and Between Groups Judging 25 Telephone Booths*

	Within Group			Between Group		
	r_{Teacher} (n = 10)	$r_{\text{IDE Student}}$ (n = 12)	$r_{\text{Ma th Student}}$ (n = 12)	$r_{\text{Teacher/ IDE Student}}$	$r_{\text{Teacher/Math Student}}$	$r_{\text{IDE Student/ Math Student}}$
Creativity	.33	.36	.45	.28	.30	.31
Prototypical Value	.33	.24	.46*	.34	.34	.41
Attractiveness	.40	.34	.38	.38	.34	.28
Interest	.37	.33	.38	.32	.34	.30
Technical Quality	.33	.25	.21	.21	.26	.23
Expressiveness	.36	.30	.34	.35	.32	.28
Integrative Capacity	.42	.19	.40	.30	.27	.16

*$p < .05$

Downloaded By: [Massachusetts Institute of Technology] At: 07:11 18 October 2009

three groups of judges. The tables show that the interrater reliability is relatively low. On most criteria the agreement among the experts is no greater than among the nonexperts. There is a difference between the two design types, with, on the whole, a higher range of correlations for the telephone booths. The relatively small variation between the 19 computer cabinet models, compared to the telephone booths, probably accounts for the lower level of agreement. Among all the groups of judges agreement on creativity is slightly higher than on most other criteria.

The between-group correlations show the same differences between the two designs as the within-group correlations, although most values are slightly lower.

Interrater reliability, using Cronbach's alpha (Winer, 1971), is presented in Table 3. These values have been calculated to enable comparison with Amabile's (1982) studies.

In Table 4 the degree of difficulty in using an assessment criterion is presented separately for each group of judges, by means of average scores and standard deviations. Again, a 10-point scale was used, a low number indicating that judging was less difficult. Judging designs on creativity seems relatively easy, suggesting that each judge has an understanding of the concept. On the other hand, judges appear to have difficulty defining technical quality, expressiveness, and integrating capacity.

Validity. Mean Pearson product–moment correlations between the different criteria are presented in Table 5. As the coefficients show, creativity is most closely related to attractiveness and interest, whereas its correlation with prototypical value and technical quality is, on the whole, small. In the case of the design teachers, nearly all correlations between the criteria are high, particularly when judging the models of a telephone booth.

The criteria "expressiveness" and "integrating capacity" give a rather obscure picture. Each group of judges seems to interpret these criteria in a different way.

To indicate the structure that underlies the assessment criteria, a principal components analysis with varimax rotation was performed on the mean scores for each group. For each separate group of judges and type of product any of the performed analyses yield a two-factor solution. Factor loadings are presented in Table 6. One of the two factors is

Table 3. *Interrater Reliability Within Groups of Judges: Cronbach's Coefficient Alpha*

	Computer Cabinet			Telephone Booth		
	IDE Teachers	IDE Students	Math Students	IDE Teachers	IDE Students	Math Students
Creativity	.79	.81	.79	.79	.86	.88
Prototypical Value	.71	.79	.74	.83	.80	.91
Attractiveness	.66	.66	.72	.86	.84	.89
Interest	.68	.76	.65	.84	.83	.87
Technical Quality	.74	.79	.84	.84	.81	.74
Expressiveness	.66	.48	.76	.84	.81	.86
Integrative Capacity	.72	.24	.77	.87	.74	.88

Table 4. *Average Degree of Difficulty in Judging Designs on Various Criteria*

	IDE Teachers		IDE Students		Math Students	
	M	*SD*	*M*	*SD*	*M*	*SD*
Creativity	3.90	3.41	3.08	1.83	2.00	1.65
Prototypical Value	4.00	3.16	5.00	3.30	4.25	2.14
Attractiveness	2.90	3.14	2.00	2.92	1.33	1.89
Interest	3.20	3.39	3.00	2.56	1.92	1.62
Technical Quality	5.00	3.16	5.42	2.97	5.50	2.68
Expressiveness	5.30	3.86	6.58	2.84	7.75	2.18
Integrative Capacity	4.90	3.90	7.67	2.10	5.83	3.24

Note: Scale values ranged from 0 to 10.

Downloaded By: [Massachusetts Institute of Technology] At: 07:11 18 October 2009

Table 5. *Mean Pearson Correlations Between Assessment Criteria, for Computer Cabinets (Below Diagonal) and Telephone Booth (Above Diagonal).*

Attribute		Creativity	Prototypical Value	Attractiveness	Interest	Technical Quality	Expressiveness	Integrative Capacity
Creativity	A	—	.60**	.89**	.92**	.72**	.84**	.84**
	B	—	.24	.76**	.89**	.14	.66**	.70**
	C	—	.36	.57**	.84**	.18	.45*	.40*
Prototypical Value	A	.33	—	.84**	.80**	.84**	.88**.	.85**
	B	–.28	—	.63**	.44*	.36	.69**	.58**
	C	–.07	—	.78**	.67**	.50**	.82**	.86**
Attractiveness	A	.77**	.70**	—	.96**	.86**	.96**	.91**
	B	.72**	.02	—	.90**	.61**	.92**	.92**
	C	.24	.72**	—	.85**	.66**	.95**	.83**
Interest	A	.90**	.50*	.89**	—	.85**	.95**	.95**
	B	.89**	.00	.80**	—	.39	.81**	.82**
	C	.64**	.34	.68**	—	.47*	.77**	.68**
Technical Quality	A	.07	.71**	.55*	.25	—	.87**	.88**
	B	–.35	.27	.04	–.38	—	.58**	.62**
	C	–.34	.67**	.63**	.27	—	.71**	.55**
Expressiveness	A	.71**	.56*	.87**	.86**	.51*	—	.95**
	B	.45*	–.07	.77**	.45*	.35	—	.86**
	C	–.13	.72**	.84**	.45*	.73**	—	.88**
Integrative Capacity	A	.67**	.72**	.92**	.80**	.72**	.91**	—
	B	.54*	.11	.65**	.60**	.16	.74**	—
	C	–.08	.75**	.83**	.47*	.62**	.89**	—

Note: A = IDE teachers; B = IDE students; C = math students.
*$p < .05$. **$p < .01$.

Table 6. *Factor Loadings of Criteria for Three Groups of Judges*

	IDE Teachers		IDE Students		Math Students	
Attribute	**Factor I**	**Factor II**	**Factor I**	**Factor II**	**Factor I**	**Factor II**
Computer Cabinet						
Creativity	.96	.01	.79	–.53	.95	–.21
Prototypical Value	.31	.83	.03	.54	.05	.86
Attractiveness	.81	.53	.93	–.02	.44	.85
Interest	.96	.22	.83	–.44	.84	.42
Technical Quality	.05	.96	.06	.91	–.19	.85
Expressiveness	.80	.48	.84	.33	.08	.94
Integrative Capacity	.70	.69	.85	.20	.14	.91
Telephone Booth						
Creativity	.94	.32	.98	.02	.97	.11
Prototypical Value	.32	.93	.23	.75	.31	.83
Attractiveness	.73	.65	.76	.63	.51	.82
Interest	.80	.59	.91	.33	.82	.5
Technical Quality	.49	.80	.08	.85	.07	.82
Expressiveness	.67	.71	.66	.68	.38	.90
Integrative Capacity	.66	.71	.70	.63	.33	.85

Downloaded By: [Massachusetts Institute of Technology] At: 07:11 18 October 2009

dominated by the aesthetic component, whereas the other factor is determined by technical quality. The factor structures for design teachers and IDE students are most closely associated. Math students show a somewhat different structure mainly through the position of the attractiveness criterion in relation to that of prototypical value. The theory that the attractiveness of a stimulus increases the more it resembles the prototypical representation of that stimulus (Purcell, 1984) seems to apply only to the math students.

Conclusions

The mean Pearson product–moment correlations within each group of judges show that interrater agreement between experts is no higher than between nonexperts or those with intermediate expertise. The results of this study with respect to the aesthetic preference for the designs do not differ from studies on artworks. Our expectations that design work judgments should show higher agreement than artwork judgments were not confirmed. However, the reliability of creativity ratings, in terms of internal consistency (Cronbach's alpha), are quite high and similar to the results of Amabile's (1982) studies. However, most studies in this field claim that alpha gives an overestimation of reliability.

The expert judges—the design teachers—did not show a higher level of agreement than the nonexpert judges—the math students. Again, the findings of a number of studies dealing with the idiosyncratic standards of experts are confirmed. On the other hand, there are marked differences in value judgments between the three groups. Design teachers show substantial correlations between the assessment criteria, much higher than among the other groups of judges. These results confirm the conclusion of Getzels and Csikszentmihalyi (1976) and Hekkert and van Wieringen (1996) that experts do not distinguish in detail between the different criteria.

Regarding the validity of creativity measurement, the relationship of creativity with other criteria is of interest. Factor analysis makes clear that technical quality and creativity are two separate factors. There is a close relationship between creativity and the two aesthetic criteria, interest and attractiveness. The relationship between creativity and prototypical value is also of interest. As was assumed by our theoretical assumption based on Purcell (1984), the results show that designs judged creative are far from being prototypical representations. The second assumption, that creative products will thus be less attractive because of their considerable divergence from the prototype, only applies to the math students.

The ease with which judges used the creativity criterion when assessing models made by first-year students can also count as a valid measurement in that we can, at least, expect face validity.

Experiment 2

The second experiment focuses on further elaboration of the results from Experiment 1. By replicating Experiment 1, we can examine the reliability of the assessment techniques and find the relationship between creativity and other criteria tested. However, the method followed in Experiment 2 differed in two respects. First, two assessment criteria, expressiveness and integrative capacity, were removed, because of their doubtful reliability and validity. Second, only one group of judges, senior IDE students, were selected. Compared to the IDE teachers, this group proved equal to the judgment task. Moreover, they were better able than the IDE teachers to differentiate between the various criteria.

A second objective of Experiment 2 was to validate the assessment technique by investigating the "nomological network" regarding the creativity concept (De Groot, 1961). Efforts were made to identify which features of the design contributed to its creativity. For this purpose judges were interviewed and also asked to fill in a "semantic scale" for a number of products. Besemer and O'Quin (1986) developed a Creative Product Semantic Scale (CPSS) based on a theoretical model in which three conceptual criteria are proposed: "Novelty (newness of processes, materials and design), Resolution (functionality, usefulness, workableness of the product), and Elaboration and Synthesis (the stylistic criteria of the finished product)" (p. 114). For each of the criteria a great number of bipolar subscales were developed, 70 in total. This CPSS was used in the study.

Downloaded By: [Massachusetts Institute of Technology] At: 07:11 18 October 2009

Downloaded By: [Massachusetts Institute of Technology] At: 07:11 18 October 2009

Method

Students and judges. Design models were selected from three different projects from the 1987 first-year design course of IDE. The first model was again a cardboard telephone booth. The second model was a cardboard shop window display. The third and final product was a drill holder, represented by technical and presentation drawings. Slides of the models and the drawings were used in the experiment.

The judges were 10 senior male IDE students, selected on the basis of having had high marks for all design work in the preceding years. They were not involved in the first experiment.

Questionnaire. A translated version of Besemer and O'Quin's Creative Product Semantic Scale (CPSS) was used in this study. The Dutch version of the scale consisted of a list of 69 bipolar 7-point subscales. For every product assessed the participant filled in the whole list.

Procedure. The judgment task was carried out in two sessions. In the first session 55 models of the telephone booth were judged. The assessment criteria were: creativity, prototypical value, attractiveness, interest, and technical quality. The instructions were the same as those used in Experiment 1, except that the concept of prototypicality was no longer defined by the words "goodness of example" but by "prototypical value." Also, the measurement was done in a different way. Before the design products were presented, each judge was asked to sketch his prototype of a telephone booth on a piece of blank A4 paper. By this method it was anticipated that a more reliable reproduction of the judges' prototype could be determined.

All 55 designs were then judged on prototypical value. Two slides of each design from different camera angles were projected side by side for 8 sec. The same procedure was repeated for the other criteria. The criteria were assessed in a fixed sequence: prototypical value, creativity, technical quality, and attractiveness. The random sequence of slides was changed every time the slides were assessed on another criterion. Each judge undertook the task individually, rating the designs on a 10-point scale. The instructions were also the same as in Experiment 1.

In a second session, 10 months later, the same judges were asked to perform four assessment tasks:

- Assessment of the 35 shop window displays on three criteria: creativity, technical quality, and attractiveness. The sequence of the criteria was fixed, but that of the slides varied among participants.
- A semi pair-comparison analysis of two highly creative designs and two designs with a low creativity rating from the first session (telephone booth), by means of an interview with a judge. While the slides were shown side by side, judges were asked to give a detailed explanation of why one was creative and the other was not. This procedure was then repeated with another pair of products. The experimenter only interrupted the judge to summarize what had been said. The protocols were recorded on tape and afterward reproduced completely in a verbal report.
- Assessment of the 30 drill holders on three criteria: creativity, technical quality, and attractiveness. The designs were presented on slides, two for each design: a rendering (presentation drawing) of the design and a technical drawing. This technical drawing was meant to improve the assessment of technical quality and thus raise the reliability of this criterion.
- The same four designs used in the pair-comparison task were judged on the CPSS, the aforementioned semantic scale. Each participant was shown the first slide and asked to fill in the scale. After completion of the scale, the next slide was presented and the scale was filled in again. This procedure was repeated for all four designs.

Results

Reliability. The level of agreement between judges is defined by calculating mean Pearson product–moment correlations and Cronbach's alpha. Results of the judgments of telephone units, displays, and drill holders are presented in Table 7.

Table 7. *Mean Pearson Correlations Between Judges, and Cronbach's Coefficient α*

	Telephone Booth (k = 55)		Display (k = 35)		Drill Holder (k = 30)	
	r	α	r	α	r	α
Creativity	.48**	.89	.57**	.93	.49**	.90
Prototypical Value	.18	.69	—	—	—	—
Attractiveness	.38**	.86	.42*	.88	.41*	.85
Interest	.44**	.89	—	—	—	—
Technical Quality	.33*	.84	.28	.80	.34	.82

* $p < .05$. ** $p < .01$.

Downloaded By: [Massachusetts Institute of Technology] At: 07:11 18 October 2009

In all three measurements the mean correlation for creativity is higher than for any other criterion. In comparison with Experiment 1, the interrater agreement is substantially improved. Consequently, the reliability of the creativity measurement, as defined by coefficient α, is high. The low agreement on prototypical value could be expected because of the individual differences judges showed in their prototypical representation of a telephone unit.

Validity. Looking at the relationship between the various criteria, creativity is again closely related to interest and attractiveness, and it correlates poorly with prototypical value and technical quality.

In Tables 8 and 9, correlations between criteria are presented for the three designs measured.

The results confirm the findings of the first experiment. The assumption concerning the negative or zero correlation between creativity and prototypical value is supported.

The factor structure, obtained through principal components analysis, looks similar to that in Experiment 1, showing an aesthetic factor and a technical quality factor. The factor loadings obtained after varimax rotation are presented in Table 10.

Oral judgment. In individual interviews the judges were asked to explain the differences between the pairs of highly creative and less creative products (telephone booths). Because of the agreement between the judges' statements it was possible to classify them into five separate categories. A selection of encoded data is presented in Table 11. The five categories include:

- Expectation pattern: When presented with an unknown design, judges apparently compare it with a representation of the product they already have in mind. Unexpected designs or design elements have a greater chance of being assessed as creative.
- Integration of various relevant criteria: The added value of the design is dependent on successful synthesis. In designs of low creativity, components are stuck together without any feeling for interdependence or the "wholeness" of the design. Imagery is an important aspect of integration.
- Form and function: Creativity should not only be expressed in general terms but also through specific characteristics: through associations, dynamic expression, use of colors and materials, and balancing details that serve both functional and aesthetic properties.
- Impact on the observer: A creative design triggers attention and fantasy and acts on our emotions.
- Commitment of the designer (Weisberg, 1988): The extent to which the designer felt challenged by the design task. This attitude is demonstrated by the designer who actively searches for new ideas and forms, confronts difficulties, and takes risks.

Most of the categories also apply to any domain, but the integration category together with the form and function category are domain-specific.

CPSS. The same designs—two with high and two with low creativity ratings—were assessed by the

Table 8. *Mean Correlations Between Criteria in Judging Telephone Booths*

	Telephone Booth				
	Creativity	**Prototypical Value**	**Attractiveness**	**Interest**	**Technical Quality**
Creativity	—				
Prototypical Value	–.04	—			
Attractiveness	.70**	.16	—		
Interest	.86**	.09	.87**	—	
Technical Quality	.05	.34*	.56**	.33*	—

*$p < .05$. **$p < .01$.

Table 9. *Correlations Between Criteria, for the Display and Drill Holder*

	Display			Drill Holder		
	Creativity	**Attractiveness**	**Technical Quality**	**Creativity**	**Attractiveness**	**Technical Quality**
Creativity	—					
Attractiveness	.87*	—		.87*	—	
Technical Quality	–.04	.32	—	.79*	.83*	—

* $p < .01$.

Table 10. *Factor Loadings of Assessment Criteria for Three Design Types*

	Telephone Booth		Shop Display		Drill Holder	
	Factor I	**Factor II**	**Factor I**	**Factor II**	**Factor I**	**Factor II**
Creativity	.93	–.14	.98	–.11	89	.42
Prototypical Value	–.07	.81	—	—	—	—
Attractiveness	.88	.37	.95	.28	.78	.57
Interest	.96	.14	—	—	—	—
Technical Quality	.26	.81	.06	.99	.46	.89

judges using the CPSS. The scores on the bipolar subscales are presented in Table 12.

The novelty dimension—with the factors original, surprising, and germinal—discriminates very well between the high and low creative designs on all subscales. Regarding the dimensions resolution and elaboration/synthesis, only some factors vary with the level of creativity: valuable, elegant—except for the subscale harmonious–jarring—and complex. The subscales sufficiency and meaningfulness also discriminate between the two pairs of designs and differ in this respect from the other subscales with which they form a factor. On the factors understandable and useful, the results are not very clear, although one might expect a positive score for the designs with a high creativity rating.

Conclusions

A higher level of agreement seemed to be reached with the selection of a more homogeneous group of judges; that is, senior male students from the same year who had high design marks. The mean correlations between judges were higher than in Experiment 1. Again, there is a close relationship between creativity and the aesthetic criteria attractiveness and interest. Nevertheless, the assumption that creativity is a separate construct that can be distinguished from adjacent constructs is adhered to. Sketching one's own internal representation before viewing the series of designs to be judged was a useful way of measuring the influence this representation had in relation to

Downloaded By: [Massachusetts Institute of Technology] At: 07:11 18 October 2009

Table 11. *Oral Statements by 10 Judges to Designs with High and Low Creativity Ratings*

High Product Creativity	Low Product Creativity
Expectation pattern (k = 11) • The designer made a great stride forward. • At first sight the design is unknown, different from existing solutions. • The designer let himself go from the box idea. • The point of reference plays an important role.	Expectation pattern (k = 16) • The design comes up to one's expectations: a box in which to phone. • It cannot be distinguished from existing booths. It does not add anything. • When you imagine a telephone booth, you think of a heavy construction. • A straightforward design: solid where you expect it, transparent where you expect it. • Originated from an old-fashioned, conventional train of thought.
Integration of various dimensions (k = 16) • The design gives the feeling that the designer reflected on the function in relation to the implications of using a telephone. It asks for imagination. • Only if elements are integrated (shape, function, association, color, situation), an original idea possible. • The structure determines the creativity of the design in that the concept itself must be good. • The choice of a tube frame together with transparent, synthetic material yields a very good match.	Integration of various dimensions (k = 12) • All elements are there: a telephone, a top, a seat. It will function very well, but it is not creative. Everything is simply stuck together. • Shapes are still unconnected. The elements must be integrated. In other words, the design appears still in its infancy. • The function of this enclosure, of the color is unclear. The designer did not reflect upon the act of telephoning. • Here the designer chose very thick material, while at the same time he realized the booth had to be made transparent by using glass. The materials are not combined effectively.
Form and function (k = 29) • The designer has to draw a cylinder three-dimensionally, or to manipulate simple models of it. • The open structure together with the round shape is appealing. The design is fragile, refined. The color corresponds to modern design. • The material is exploited fully. Tubes are partly curved, partly not; designed not only for the purpose of support, but also for aesthetic embellishment. • Form details are balanced. The emphasis on some of these details is chosen deliberately. • The visual emphasis is toward the top. Hence, the design appears more dynamic. • The designer played with forms with space by applying transparent areas. It seems to float through space.	Form and function (k = 18) • The square shape is two-dimensional. It is a "drawing-table design." • When people seek solutions for protection, they build a square, a box. Working with straight lines is the easiest way. Thinking in terms of a box is not very creative. • The material is hardly exploited. The designer applied standard elements. It remains a semi-manufactured article. • The visual emphasis of this booth is toward the ground, due to the design's lower half solidity. It is obvious that one board up the bottom. Even the glass elements cannot change the static impression.
Impact on the observer (k = 27) • The designer added value to the design by deliberately using associations. For example big shells for the earpiece corresponding to a part of the activity. • The design stimulates fantasy. It evokes tension and emotion because of its unexpected image. It holds attention for a while.	Impact on the observer (k = 17) • The design suggests a dug-out, a tram-shelter, the entrance of a big building, but it does not speak the language of the telephone and of the activity itself. • The design does not fascinate. It is boring, like looking at a white wall. Among so many visual impressions this is one you immediately forget.
Designer's commitment (k = 7) • The designer searched for new forms. He spares no pains to come up with an attractive design. He takes risks in that it is uncertain if the result is functional and satisfies the client. • The designer has considered the problem of being original, the technical ingredients given; he is facing the difficulties like the connection of a curved surface on a cylinder.	Designer's commitment (k = 7) • The designer keeps on the easiest side. • Both in the concept and detail, nothing is searched for. Only minimum requirements are met. • The designer only tried to polish the primitive form, which is not creative.

Note: k = total number of responses within the category.

Downloaded By: [Massachusetts Institute of Technology] At: 07:11 18 October 2009

Table 12. *Means and Standard Deviations on the Creative Product Semantic Scale (CPSS) Judging Designs with Low and High Product Creativity Ratings (PCR)*

	Product 1 Low PCR		Product 2 Low PCR		Product 3 High PCR		Product 4 High PCR	
	M	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*
Novelty								
Original								
Exciting–Dull	6.50	.53	6.70	.48	1.80	.63	2.10	.57
Zippy–Bland	6.40	.70	6.60	.70	1.90	.74	2.10	1.10
Fresh–Overused	6.90	.32	6.30	.95	1.50	.53	1.40	.52
Eccentric–Conventional	6.80	.42	6.60	.70	2.20	.92	1.60	1.07
New–Old	6.30	.48	5.90	1.20	1.50	.71	2.00	1.15
Novel–Predictable	6.90	.32	6.00	1.49	2.10	.57	1.90	.57
Unusual–Usual	6.60	.52	5.50	1.84	1.80	.63	1.80	.63
Unique–Ordinary	6.70	.67	6.20	.92	1.80	.63	2.00	.67
Original–Commonplace	6.70	.67	6.70	.48	1.60	.70	1.40	.52
Surprising								
Startling–Stale	6.50	.97	6.40	.97	1.90	.74	2.00	.82
Surprising–Customary	6.60	.70	6.00	1.56	1.80	.92	1.60	.70
Astonishing–Commonplace	6.80	.42	6.70	.48	2.50	.85	2.10	.74
Astounding–Common	6.70	.95	6.80	.42	3.40	.97	2.90	.74
Shocking–Ordinary	6.60	.97	6.40	.70	3.30	.67	2.40	.70
Unexpected–Expected	6.10	1.45	5.70	2.11	2.50	1.08	1.90	.74
Germinal								
Trendsetting–Warmed Over	6.80	.42	6.10	1.20	2.00	.82	2.30	1.49
Revolutionary–Average	6.60	.97	6.50	1.08	2.50	.85	2.60	1.65
Radical–Old Hat	6.80	.42	6.40	.70	2.10	.74	2.10	.74
Resolution								
Valuable								
Priceless–Worthless	5.70	1.16	5.90	1.45	3.20	.79	3.50	1.08
Valuable–Worthless	5.50	.97	5.80	1.40	2.20	.42	3.30	1.42
Important–Unimportant	6.50	1.27	5.90	1.45	3.10	.88	3.40	1.17
Significant–Insignificant	6.20	1.03	6.40	1.07	2.60	.84	3.10	.99
Essential–Inessential	5.40	1.17	5.90	1.45	2.70	1.06	3.90	1.37
Necessary–Unnecessary	5.40	1.07	5.70	1.77	2.90	.99	4.60	1.51
Logical								
Logical–Illogical	2.70	1.34	3.80	2.62	3.00	1.49	4.10	1.66
Makes Sense–Senseless	4.90	1.34	5.50	1.58	2.30	1.49	4.00	1.49
Correct–Incorrect	4.30	2.16	4.40	1.84	2.70	.67	3.60	1.35
Appropriate–Inappropriate	3.90	1.97	4.90	2.08	2.20	.63	3.50	1.51
Adequate–Inadequate	3.20	1.48	4.10	1.52	3.10	1.60	3.70	1.49
Useful								
Effective–Ineffective	3.20	1.48	4.00	2.49	2.50	1.18	3.70	1.57
Functional–Nonfunctional	2.20	.92	3.20	1.99	2.70	1.42	4.00	1.56
Feasible–Unfeasable	2.40	.70	3.90	2.18	2.00	.82	3.70	1.64
Operable–Inoperable	1.40	.70	1.60	.52	2.60	.97	2.60	1.07
Useful–Useless	4.00	1.56	5.20	1.99	2.40	.84	3.70	1.57
Workable–Unworkable	2.70	1.25	2.80	1.32	2.60	1.26	3.30	1.06
Usable–Unusable	2.30	.67	3.50	2.27	2.20	1.23	3.80	1.69
Durable–Flimsy	1.90	.99	2.60	1.26	5.20	1.32	3.50	1.27
Substantial–Insubstantial	4.30	1.64	4.60	2.01	2.60	.97	3.60	1.90

(*continued*)

Downloaded By: [Massachusetts Institute of Technology] At: 07:11 18 October 2009

Table 12. *Continued*

	Product 1 Low PCR		Product 2 Low PCR		Product 3 High PCR		Product 4 High PCR	
	M	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*
Elaboration and Synthesis								
Organic								
Orderly–Disorderly	2.20	1.23	1.60	.52	3.50	1.51	3.60	1.78
Arranged–Disarranged	2.40	1.35	1.50	.71	4.70	1.83	5.20	1.75
Organized–Disorganized	3.20	1.32	2.00	1.05	3.10	1.20	4.40	1.58
Formed–Formless	5.90	.99	6.20	1.32	2.20	.92	3.50	1.35
Complete–Incomplete	3.00	1.49	3.10	1.45	2.80	1.55	3.20	1.14
Whole–Partial	2.30	1.16	2.50	1.72	3.00	2.05	3.40	2.12
Sufficient–Insufficient	5.50	1.51	5.30	2.00	1.30	.48	3.20	1.40
Perfect–Imperfect	5.70	1.16	6.20	1.23	3.00	1.25	4.70	1.16
Elegant								
Harmonious–Jarring	3.60	.97	3.40	1.65	2.80	1.96	3.10	1.73
Graceful–Awkward	6.20	.79	5.70	1.57	1.80	.79	3.60	1.43
Charming–Repelling	6.00	.67	5.80	1.14	1.70	.67	2.60	.97
Elegant–Coarse	6.00	.82	5.10	1.60	1.60	.70	3.70	1.57
Attractive–Unattractive	6.30	.95	6.60	.70	1.60	.70	2.80	1.48
Complex								
Intricate–Simple	5.90	1.60	5.80	1.23	3.90	1.10	3.80	1.14
Complex–Simple	6.00	1.05	5.70	1.95	4.10	1.10	3.50	1.58
Ornate–Plain	6.50	.53	6.80	.63	1.90	.74	3.50	1.65
Complicated–Straightforward	6.20	1.14	5.50	1.78	3.70	1.16	3.40	1.17
Interesting–Boring	6.80	.42	6.70	.67	1.70	.82	2.40	.70
Understandable								
Meaningful–Meaningless	6.30	.95	6.40	1.07	1.90	.74	3.40	1.07
Understandable–Mysterious	1.90	.99	2.40	1.65	3.80	1.75	3.90	1.20
Intelligible–Unintelligible	2.60	1.35	3.40	2.17	2.30	.82	3.30	1.42
Clear–Ambiguous	3.30	1.89	4.20	2.15	2.30	1.57	3.10	1.60
Explicit–Implicit	3.50	1.84	3.40	2.01	2.20	.79	2.70	1.77
Self-Explanatory–Unexplained	2.40	1.71	4.00	2.36	2.80	1.62	5.20	1.14
Well-Crafted								
Skillful–Bungling	5.30	1.42	4.80	1.48	2.00	1.05	3.90	1.73
Well-Made–Botched	5.10	1.52	4.40	1.84	2.10	.88	5.00	1.56
Well-Crafted–Crude	5.50	1.35	5.00	1.33	2.70	1.25	4.50	1.51
Meticulous–Sloppy	4.90	1.66	4.00	1.56	3.10	1.52	5.00	1.41
Skilled–Unskilled	5.30	1.57	5.50	1.65	2.60	.97	4.20	1.23
Expert–Inept	6.60	.70	6.10	1.45	2.80	1.40	4.50	1.58
Careful–Careless	5.70	1.57	4.80	2.15	1.70	.67	4.70	1.42

Downloaded By: [Massachusetts Institute of Technology] At: 07:11 18 October 2009

other criteria. When people judge the creativity of products, they really seem to compare the object to be judged with their internal object representation, their prototype. As was expected, the correlation between prototypical value and creativity was found to be low in this experiment, whereas the correlation between this value and attractiveness was higher.

In two of the three sessions in Experiment 2 the correlation between creativity and technical quality was low, as was found in Experiment 1. However, in judging drill holders this correlation was unexpectedly higher. In this particular session, the technical drawings of the designs were also presented on slides. It follows then that when offered explicit technical information the judge may be better able to assess the technical quality of the product, and thus its impact on creativity. Nevertheless, even in this session the agreement on this criterion was low.

The pair-comparison of high and low creative designs provided extra information on what is understood by creativity. The aspects mentioned form a mixture of domain-specific and general elements.

Downloaded By: [Massachusetts Institute of Technology] At: 07:11 18 October 2009

First, the expectation pattern the judges had was mentioned by all of them. It confirms the finding that people use their internal representation of the object as a frame of reference. Second, the extent to which the designer succeeds in integrating different design dimensions, such as form, function, and use, is manifested in the design itself. Styling also seems to contribute further to the creativity of the design. Judges identify the use of certain techniques by which objects look dynamic, fragile, and balanced, three adjectives with a positive bias. The impact of the design on the observer, a third aspect, can thus be manipulated by the designer if he is skilled in using such techniques. The final aspect mentioned is the commitment of the designer; it seems astonishing that judges can deduce this simply by looking at the designs. All these aspects show that judges are very capable of defining what they understand by creativity. Moreover, the fact that there is general agreement in their definitions shows that creativity differs from related concepts. The results on the CPSS show that designs with a high creativity rating are judged to be original, surprising, germinal, valuable, and elegant.

General Discussion

As long as no absolute criterion of creativity exists, the assessment of creativity remains dependent on subjective judgment. In design education, and in other design selection procedures, it would be helpful to rely on expert judges, because that is what happens at present. But in the area of art it has become clear that the judgment of experts is not very reliable, nor does it differ substantially from that of the nonexperts. We might expect a difference in the design domain, because, unlike art, design includes more objective aspects that mainly involve the functionality and technical quality of the design. In this study, however, judgments by people with different levels of design expertise show that, in this respect, design assessment does not differ from art assessment. Interrater agreement between design teachers with professional design experience is low, and experts judge no better than nonexperts. They are even less able to differentiate between the assessment attributes than judges with an intermediate level of expertise (design students). This is in line with art assessment studies (Getzels & Csikszentmihalyi, 1976; Hekkert & van Wieringen, 1996). Studies in other areas, however, are comparable to art and design. An example is the assessment of the quality of grant applications concerning scientific research in The Netherlands. The mean product–moment correlation among the five judges, based on the final judgment, was only .14 (Hofstee, 1983). When looking at the separate judgment criteria the mean correlations were between .12 and .29. This example demonstrates that, even when large sums of money are invested and judges are doing a rather poor job, confidence in human judgment remains.

If the objective is to make judgments more reliable and valid, then one way is to select a homogeneous group of judges. In our second experiment judges were senior male students from the same year with high design marks, and their agreement was substantially better than in the first experiment. The assumption was that similar cultural background and learning experiences will produce similar design knowledge and opinions. However, such a selection is hardly practicable in everyday situations.

Because of the low level of agreement between judges, the validity of the creativity measurement in this study is also questionable. However, validity can also be studied by analyzing the nomological network between adjacent concepts (De Groot, 1961). Both experiments showed that creativity is closely linked to aesthetic criteria such as attractiveness and interest, whereas a correlation with the technical quality and people's prototypical (mental) representation of the design is much lower. Nevertheless it is argued that creativity can be considered to be a separate concept. This statement is supported by the results of the interviews in Experiment 2 based on pair-comparison analysis. This method by which judges were asked to comment on pairs of designs, one of which had a high creativity rating and one with a low creativity rating, proved to be especially valuable in providing domain-specific information. When judges are asked to express in words the underlying aspects of creativity, they refer to three elements: (a) The impact on the observer: unexpectedness, emotions; (b) design characteristics regarding form and function: integration of shape, function, emotions, material, texture and color; associations between form and function; and (c) the designer's commitment as manifested by his willingness to search for new solutions and to take risks. This

statement shows that elements of the design process are reflected in the design itself, as was already suggested by Weisberg (1988).

Finally, the results of the responses to the CPSS of Besemer and O'Quin (1986) showed that in judging highly creative designs, the novelty dimension, including originality and surprise, was more respected by the judges than any other dimension. It was surprising that the factor usefulness seemed not to be important in discriminating between designs with high and low creativity ratings. This aspect or related dimensions such as appropriateness is always included in definitions of creativity. The result of this study confirms a study by Runco and Charles (1993) with respect to the importance of appropriateness in judging creative artworks.

Downloaded By: [Massachusetts Institute of Technology] At: 07:11 18 October 2009

References

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

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

Besemer, S., & O'Quin, K. (1986). Analyzing creative products: Refinement and test of a judging instrument. *The Journal of Creative Behavior, 20,* 115–126.

Child, I. L. (1970). *A comparison of art preferences in culturally varying groups.* Paper presented at the VIth International Colloquium of Empirical Aesthetics, Paris.

De Groot, A. D. (1961). *Methodology.* The Hague, The Netherlands: Mouton.

Getzels, J., & Csikszentmihalyi, M. (1976). *The creative vision: A longitudinal study of problem-finding in art.* New York: Wiley-Interscience.

Gordon, D. A. (1956). Individual differences in the evaluation of art and the nature of art standards. *Journal of Educational Research, 50,* 17–30.

Hekkert, P., & Wieringen, P. C. W. van. (1996). Beauty in the eye of expert and nonexpert beholders: A study in the appraisal of art. *American Journal of Psychology, 109*, 389–407.

Hofstee, W. K. B. (1983). Rating of requests for grants in educational research: A psychometric evaluation. *Journal of Educational Research (Tijdschrift voor Onderwijsresearch), 8,* 273–284.

Hofstee, W. K. B. (1985). Rather clinical. Limits on the objectivity principle in assessment and selection. *Dutch Journal of Psychology (Nederlands Tijdschrift voor de Psychologie), 40,* 459–473.

Purcell, A. T. (1984). The aesthetic experience and mundane reality. In W. R. Crozier & A. J. Chapman (Eds.), *Cognitive processes in the perception of art* (pp. 189–210). Lisse, The Netherlands: Elsevier.

Runco, M. A., & Charles, R. E. (1993). Judgments of originality and appropriateness as predictors of creativity. *Personality and Individual Differences, 15*, 537–546.

Runco, M. A., McCarthy, K. A., & Svenson, E. (1993). Judgments of the creativity of artwork from students and professional artists. *The Journal of Psychology, 128*, 23–31.

Temme, J. E. V. (1983). *There is disputing about taste.* Unpublished doctoral dissertation, Utrecht University, Utrecht, The Netherlands.

Trowbridge, N., & Charles, D. C. (1966). Creativity in art students. *The Journal of Genetic Psychology, 109,* 281–289.

Ward, W. C., & Cox, P. W. (1974). A field study of nonverbal creativity. *Journal of Personality, 42,* 202–219.

Weisberg, R. W. (1988). Problem solving and creativity. In R. J. Sternberg (Ed.), *The nature of creativity* (pp. 148–176). New York: Cambridge University Press.

Winer, B. J. (1971). *Statistical principles in experimental design.* New York: McGraw-Hill.