Comparing the Creativity of Children’s Design Solutions Based on Expert Assessment

Binh Thang
VU University Amsterdam
Faculty of Economics and Business Administration
De Boelelaan 1105
1081 HV Amsterdam
The Netherlands
htg600@few.vu.nl

Wouter Sluis–Thiescheffer
Tilde Bekker
Berry Eggen
Eindhoven University of Technology
Faculty of Industrial Design
Den Dolech 2
5600 MB Eindhoven
The Netherlands
{r.j.w.sluis; m.m.bekker; j.h.eggen}@tue.nl

Arnold Vermeeren
Huib de Ridder
Delft University of Technology
Faculty of Industrial Design Engineering
Landbergstraat 15
2628 CE Delft
The Netherlands
{a.p.o.s.vermeeren; h..deridder}@tudelft.nl

ABSTRACT
This paper describes a study in which the outcome of early design sessions with eight-to-twelve-year old children is assessed through expert judgment. Experts compare the outcomes of two early design methods: brainstorming and prototyping. The design case was to come up with a solution for incapacitated children that need to attend class from home. The hypothesis is that children will generate more creative design solutions when prototyping than when brainstorming, because we reason that prototyping requires a wider range of intelligences according to Gardner’s Theory of Multiple Intelligences. The outcome of the sessions is assessed on creativity and five explanatory criteria. The results show that a brainstorming method generates design solutions that are more creative. However, both methods produce creative design solutions; the brainstorming sessions generate more surprising and novel design solutions, the prototyping results are considered more relevant and workable.

ACM Classification Keywords
D.2.10 [Software]: Design – Methodologies.
H.5.2 [Information Interfaces and Presentation]: User Interfaces – User-centered design.
J.5.2 [Computer Applications]: Social and Behavioral Sciences – Psychology.

Author Keywords
User-centred design; methods; children; creativity assessment

INTRODUCTION
How creative and innovative is the contribution of children in a participatory design process? We present a study in which we determine the quality of the outcome of design sessions with children through expert judgment. Examining the actual contribution of children to the design process is an important topic in the area of interaction design and children [10,18].

Previous work describing children’s contribution to early, middle and later design phases used a wide range of measures to describe this contribution. For example, the study on the KidReporter method examined the quality of the output of early design methods by assessing the children’s engagement during various design activities [5]. Barendregt et al. [4] focus on a later stage in the design process; they compared the contribution of children to usability evaluations on the basis of children’s verbal skills. An example of work that compared the relevance of children’s input is presented by Kelly et al. [15]. They varied the number of design iterations with children to optimize the relevance of the relation between designers and children as co-designers. In this paper we focus on assessing the contribution of children early in the design process in terms of creativity of their design solutions as judged by experts. The rationale behind it is that the design process benefits more from the input if the designer considers the children’s design solutions creative than if the generated design solutions are found to be trivial and obvious.
Although most findings about the contribution of children to the design process have been very positive, so far no rigorous comparisons of methods have been made based on the creative value of the output of participatory design sessions have been made. To design such a study is a complex endeavor. First of all, the set-up of the design sessions should allow for a fair comparison between different types of output [17]. For example, how to compare the value of a child’s mood board with the value of a child’s acting out behavior is not obvious. Secondly, the concept of creativity can imply innovation and originality, but it can also refer to bizarre results that are not very useful. Hence the assessment of creativity should comprise both originality and usefulness. Finally, we need a representative group of design experts to value the design output. With this paper we make a first attempt to solve these issues, and we will discuss to what extend we succeeded.

In this paper we propose an approach to assess the creative products of eight-to-twelve-year-olds. The study compares the design solutions written down in a brainstorm session and the design solutions created in a prototyping session. First, we describe briefly a framework which allows us generate hypotheses about comparisons of different methods. Next, we deal with the differences between the expressive media of the design solutions. Then we will define creativity and describe the starting points for assessing the creative value of the design solutions of eight-to-twelve-year-olds. Finally, we will describe the set-up and the preliminary results of our study. The discussion section will provide directions for further research.

**COMPARING DIFFERENT EXPRESSIVE MEDIA**

The creation of the design solutions was part of a previous study [19]. In that study we compared how well the design space was explored in terms of the number of options and criteria for a design solution when brainstorming is used or when prototyping is used. In this section we describe a framework that we used in the first study to hypothesize about the exploration of the design space by children of eight-to-twelve year old. The same framework was used in the current study to hypothesize about the expected creative value of the design solutions. We also describe a group of children that generated these design solutions, and how the design solutions can be compared.

**Framework expects more from prototyping**

To reason about expected differences between methods, we have developed a framework that could explain the suitability of design methods for children based on the Theory of Multiple Intelligences of Howard Gardner [14]. The theory is used by teachers in primary schools, to think about educational methods [2]. For example, it helps them find solutions for children that are predominantly motivated to be addressed in one intelligence, but need to develop another. In this framework we consider the intelligences as skills with which design problems can be solved. The seven intelligences defined by Gardner are linguistic, logical-mathematical, spatial-visual, bodily-kinesthetic, interpersonal, intrapersonal and musical. The act of designing requires use of these intelligences in three ways:

1) To communicate
   a) i.e. understand instructions
   b) i.e. express design solutions

2) To perform using the media of the design method

3) To understand and perform within the design domain

In this study we focus on the second aspect. One of the underlying assumptions of our framework is that the more intelligences a method addresses the better the design space is explored. The design space is explored better, because the number of different intelligences required to apply a specific method determines the number of different design trajectories a child can choose to solve the design problem. Furthermore, if a child is most comfortable in activities associated with one dominant intelligence, the method involving most intelligences is most likely to address this child.

**Hypothesis**

For the previous study, design space exploration was measured in terms of the number of options and criteria for a design solution. For brainstorming, children require mainly linguistic and social skills, whereas for prototyping children also require visual-spatial and bodily-kinesthetic skills. Following the above line of reasoning, we expected that with a prototyping method, the children would explore the design space better. In analogy, for this study we expect that a better design space exploration leads to generating more ideas, including more creative ideas. Therefore we expect that design solutions generated through a prototyping session will be assessed as more creative than the solutions generated through brainstorming.

**Comparing explanations of design solutions**

It is difficult to compare brainstorming output (a piece of paper with a child’s written description of a device) with prototyping output (a 3d model of a device built from scrap materials), because the output media are so different. How to account for the difference in visual appearance? A brief description resulting from a brainstorm to a certain extent is explicit in describing required elements of a design solution, but leaves many degrees of freedom to understand how the design solutions of the children could be visualized. In prototyping, it is the other way around. To a certain extent the prototype is a visualization of the child’s design solutions, but unfortunately it is often not self explanatory at all. In both cases it is not clear what the
relative importance is of all described or crafted ideas for the design solution.

In the previous study, sixty children, nine to eleven years old, performed both a brainstorming session and a prototyping session individually. We asked the children to design a solution for attending classes online. We provided a scenario describing a child who had broken his or her leg, and was therefore immobilized. However, the child was mentally fit enough to attend class. We asked children, in groups of five, to envision devices (for use at home and in class) that would facilitate attending class with all necessary means.

Figure 1  A 10-year old girl showing her design solution

The children had to either describe (brainstorm) or create (prototyping) their design solution individually. After they finished their design solution, they were asked to explain it to the group.

To measure differences in design space exploration, the discussions were transcribed and analyzed in terms of the options and criteria the children mentioned in their explanations and the discussion. The results showed that the children mentioned significantly more options after a prototyping session than after a brainstorming session [19]. Hence we concluded that prototyping provides a better design space exploration.

To assess the quality of the design space exploration in terms of creativity, in this study experts will judge a representative subset of these sixty transcripts. The research question is whether experts will judge the explanations of prototyped design solutions as more creative?

CREATIVITY DEFINED AS ORIGINAL AND USEFUL

Creativity is strongly associated with originality. However, as Horn and Salvendy [13] point out, from a product development view a creative and successful design solution should be both original and useful. Therefore we will use their definition for creativity: “The individual or group process that results in an artefact (solution, product, thought, idea, art, etc.) that is judged as original and useful.” MacCrimmon and Wagner [16] proposed to further specify original and useful by novelty, non-obviousness, workability, relevance, and thoroughness. In our study, we will ask experts to assess the design solutions on these five characteristics.

ASSESSING CREATIVITY

To judge creativity with subjective measures, we looked for a reliable method that is based on human judgment. However, only a few creativity assessment studies have been performed in the domain of design, the assessments are more often applied in the field of art [1,6,9]. Although design work is more suitable for objective measures than art work, not many controlled experiments have been conducted that assess the creativity of design work. For our study the work of Besemer and O’Quin [7] on the creative product semantic scale (CPSS) and the work of Amabile [1] on subjective Creative Assessment Technique (CAT) has informed the manner in which we assess creativity. The work of Christiaans [9] has inspired our study set-up in elaborating on the subjective creativity assessment by combining the use of CAT and CPSS.

The Creative Product Semantic Scale

The Creative Product Semantic Scale (CPSS) developed by Besemer and O’Quin [6] is an evaluation instrument using scales based on the standard principles of semantic differentiation. CPSS measures dimensions of how well the creative product is crafted, as well as its originality and appropriateness. The Creative Product Semantic Scale (CPSS) is a validated and reliable evaluation instrument that allows untrained judges to make informed judgments of creativity in products [8]. The CPSS uses 55 adjective item pairs on a 7-point scale. The 55 items are divided into three dimensions: novelty, resolution, and style. The purpose of the CPSS is to improve judgments made by raters or evaluators, so that they carefully consider all elements of the product [7]. The CPSS is intended to be usable across domains, and by non-expert judges [6].

Although this scale is considered to be of high methodological quality, administering the scale is too time-consuming for frequent use. Using it to its full extent would put a high workload on the participants. Moreover, the results would not provide insights into what motivates the
Consensual Assessment Technique
The Consensual Assessment Technique (CAT), developed and tested by Amabile [1], is based on a consensual definition of creativity, as the assessment relies on a consensus of independent expert judges. It assumes that people recognize creativity when they see it. Thus, creativity is a subjective quality that is independently assessed by judges [13]. Amabile proposed that the most valid way to measure creativity is by using experts’ subjective assessment of creative products [12].

Application of CAT is based on three important assumptions. The first assumption is that creativity is a subjective quality that is independently assessed by judges [13]. Amabile proposed that the most valid way to measure creativity is by using experts’ subjective assessment of creative products [12].

Application of CAT is based on three important assumptions. The first assumption is that creativity is a subjective quality that is independently assessed by judges [13]. Amabile proposed that the most valid way to measure creativity is by using experts’ subjective assessment of creative products [12].

In making judgments, the judges should be selected on the basis of their expertise in a relevant domain. For this domain they have developed their own implicit criteria for creativity. Furthermore, they are to use these subjective definitions of creativity in making the ratings [11].

Although Amabile’s CAT implies a lower workload and is therefore more convenient for repetitive use, the result of such an assessment leaves open the reason why a certain product is considered more creative than another product.

Combining CPSS and CAT
Christiaans [9] combined both techniques in a study into creativity. He compared the judgments of different groups of design experts, to understand what level of expertise is most ideal in applying the CAT. He compared three groups of experts: design teachers, design practitioners and master design students with high grades. All design solutions were generated by one design method. He found that the judgments of design master students were good predictors for the judgments of design experts. Furthermore, he found that the judgments of these students were more reliable than the judgments of design teachers and practitioners.

To understand whether the judgments on creativity of one group of experts is as good as another, Christiaans had to ask the judges to elaborate on their judgments. The CAT was too limited for that purpose. He overcame the problems with CAT by utilizing CPSS as a further specification of CAT. In other words, in addition to creativity as a generative value, he used the subscales of CPSS to find explanations for the differences in creative value of the design solutions in different groups of experts. Therefore he designed a study composed of two sessions. The first session aimed to collect quantitative data: the judges had to rate forty-four designs according to seven criteria. The second session aimed to further elaborate the results of the first session. The judges were interviewed to elaborate on their motivations for their judgments.

This experiment design is an elegant solution in which the strengths of both CPSS and CAT in assessing design solutions are combined. Therefore we adopted this setup to measure the creativity of design solutions, generated in early design methods with children.

METHOD
As explained before, we expect prototyping to be more creative than brainstorming. The five criteria novelty, non-obviousness, workability, relevance and thoroughness should explain why design solutions generated through a prototyping session are more creative than design solutions generated through a brainstorming session. In this section we will provide information about the stimuli provided to the judges, the selection criteria to be used by the judges, and the procedure for eliciting information from them.

Forty design solutions to be evaluated
In the previous study, described in detail in [19], we conducted twelve focus group sessions (as explained in the previous sections). The study resulted in 120 design solutions, sixty generated in a brainstorming session, and sixty generated in a prototyping session.

The total number of 120 design solutions was reduced to forty. First of all, the influence of the 1st session on the 2nd session created some redundancy which was evident from the transcripts. Therefore we selected only the first design solution generated by a child, hence only sixty design solutions remained. Furthermore, due to the quality of the recording of some sessions, not all sixty design solutions could be isolated from the discussions. From the remaining sample we selected a set of forty design solutions, twenty generated in a brainstorming session and twenty generated in a prototyping session.

The following part of transcript is taken from a design solution from a brainstorming session. E is the experimenter, student 3 explains his idea, and the others respond to it.
In the first session, participants were first explained what the design problem was about. After they confirmed they understood the problem of attending class while being at home, they were asked to sketch a design solution themselves in ten minutes. The aim of this exercise was to have the participant empathize with the design problem. By building up empathy, we expected to create in depth awareness of the design problem, and to make them more aware of the creative aspects in a design solution.

Following the sketching exercise, the participants were given a training session to become familiar with reading a transcript and with rating the represented design solution on creativity and five descriptive criteria; novelty, surprise, workability, relevance and thoroughness. The five criteria were presented on a form with seven-point Likert scales. For each of the five criteria the negative extreme was described as “not criterion at all” and the positive extreme was described as “extremely criterion”. The participants were asked to rate each design solution immediately after reading the transcript. The session took approximately ninety minutes.

A pilot session showed that the participants had difficulties with scoring “non-obviousness” on a Likert scale that ranged from “not non-obvious at all” to “extremely non-obvious”. Therefore we replaced non-obvious with surprising, because surprising was the closest term available in CPSS. The participants had to rate between “not surprising at all” and “extremely surprising”.

Second session: motivating the ratings for the most and least creative design solutions.

Based on the outcome of session one the design solutions that were on average rated as the two most and the two least creative design solutions were selected. The selection contained the most creative and least creative design solution in prototyping and the same pair in brainstorming.

In this session the judges were asked to perform two tasks. First, the experts were asked to explain their rationale for judging the creativity of the design solutions. They were asked to read the transcripts of two design solutions – one highly creative and one low in creativity – and then had to explain in detail why they considered one design solution to be more creative than the other. This was repeated with another pair of design solutions. The pairs of solutions were combined as follows:

<table>
<thead>
<tr>
<th>Low creativity rating</th>
<th>High creativity rating</th>
</tr>
</thead>
<tbody>
<tr>
<td>Condition 1</td>
<td>Brainstorming</td>
</tr>
<tr>
<td>Condition 2</td>
<td>Brainstorming</td>
</tr>
<tr>
<td>Condition 3</td>
<td>Prototyping</td>
</tr>
<tr>
<td>Condition 4</td>
<td>Prototyping</td>
</tr>
</tbody>
</table>

Each interview contained two design solutions from each method. Experts had to either compare design solutions from conditions one and three or from two and four.

Second, after the semi-paired comparisons, the experts were asked to judge the same four design solutions used in the interview on the CPSS subscales of the dimensions novelty and resolution. Before judging the design solutions on the
CPSS scales, each expert was asked to read the transcript of the design solution again before rating it on a 7-point Likert scale. To compensate for order effects, the sequence of the subscales differed for each design solution. The sequence of design solutions was also randomized among the judges. This procedure was followed for all four design solutions.

RESULTS
Agreement
To understand the reliability of the judges, we calculated Cronbach’s alpha for each criterion in each condition. Cronbach’s alpha indicates whether the answers of a group of respondents are consistent with each other. The results are shown in table 1. The table shows that the internal reliability of the raters is fairly high for all criteria in each condition. Nearly all criteria meet the specified threshold of $\alpha=0.7$ except for novelty in the brainstorm condition and workability in the prototyping condition.

<table>
<thead>
<tr>
<th>Cronbach’s Alpha</th>
<th>Brainstorm session</th>
<th>Prototyping session</th>
</tr>
</thead>
<tbody>
<tr>
<td>Creativity</td>
<td>0.863</td>
<td>0.881</td>
</tr>
<tr>
<td>Novelty</td>
<td>0.671</td>
<td>0.844</td>
</tr>
<tr>
<td>Surprising</td>
<td>0.739</td>
<td>0.702</td>
</tr>
<tr>
<td>Workability</td>
<td>0.769</td>
<td>0.671</td>
</tr>
<tr>
<td>Relevance</td>
<td>0.745</td>
<td>0.717</td>
</tr>
<tr>
<td>Thoroughness</td>
<td>0.796</td>
<td>0.837</td>
</tr>
</tbody>
</table>

Table 1 The internal reliability of the judges’ observations for each criterion in each condition represented by Cronbach’s $\alpha$.

Results of the first session
The effect of the design method (prototyping/brainstorming) on the creativity of the generated design solutions was tested separately for the six criteria using repeated measures. The averages and their confidence intervals for the scores on each variable are shown in figure 3. The graph shows the mean ratings on a seven point Likert Scale (x-axis) as a function of the six criteria (y-axis). Markers are used to indicate the design method.

Regarding the within-subject contrasts of interest, five criteria reached the specified .05 significance level: brainstorming is more creative ($F=23.098$, $p=0.000$), novel ($F=20.414$, $p=0.000$) and surprising ($F=45.314$, $p=0.000$) than prototyping. Prototyping is more workable ($F=7.000$, $p=0.019$) and relevant ($F=30.757$, $p=0.000$) than brainstorming. The within-subject contrast for thoroughness did not reach the significance level ($F=0.006$, $p=0.941$).

Figure 3 The average scores including a .05 confidence interval on a seven point Likert scale for each criterion.

Preliminary results of the second session
Interviews
So far only the interviews of six participants have been analyzed. The semi pair-wise comparison of the design solutions shows that most judges comply with the results from the first session. The most creative design solution from the brainstorming was again considered more creative than the least creative design solutions in both brainstorming and prototyping by all judges.

The judges did not agree on whether a design solution is required to be workable. They tend to claim that a creative design solution from an early design method does not have to be realistic and workable. However, a creative design solution does have to be relevant.

The six judges explained that novelty is a necessity in a creative design solution; however, it is a subtle characteristic. Novelty can be found in new techniques or devices, but also in new combinations of old techniques, or new combinations of techniques used in an unexpected context.

Surprise is considered more important than novelty, at this stage in the design process. A creative design solution should be surprising, at least for a moment. Although in hindsight a design solution might seem obvious and logic, a creative idea is recognized by a moment of surprise.

Finally the judges think that it is important to use different methods to come up with design solutions, to maximize the exploration of the design space.

DISCUSSION
With this study we make two contributions. First, we describe how evaluation techniques used in creativity assessment for adults can be used to explain the value of the contribution of children in a design process. Our study provided useful insights in what value designing with children provides for designers in terms of creative inspiration, depending on the design method used. Furthermore, in a previous study [19], we compared the
output of children using objective measures implemented in a coding scheme. Our current work is based on subjective perceptions of designers. Therefore the current experiment provides a method to assess children’s design output with an increased ecological validity.

Unexpectedly, the design solutions generated through brainstorming are overall considered to be more creative than the design solutions generated through prototyping. From our framework of multiple intelligences, we expected that the more intelligences involved, the more design trajectories the children would find to come up with design solutions. The more design trajectories available to solve the same problem, the higher the expected chance that children would come up with more creative ideas. However, although a prototyping session delivers more information [19], we found that the quality of the design solutions is not necessarily more creative.

The unexpected result can be explained in two ways. First, our hypothesis did not take into account the process of triangulation. Prototyping allows children to approach the design problem from different angles. However, in the process of finding a solution that makes sense from all provided angles, the children will have to solve more conflicting constraints then when they only explore the problem with one or two intelligences. The constraints will withhold them from exploring the design space in the creative extremes of one intelligence.

Triangulation still explains that the prototyping method contains more information. Prototyping elicits a more thorough exploration of the constrained design space. Therefore prototyping delivers more information, albeit less surprising and novel, and therefore less creative than brainstorming.

A second explanation could be that the experts thought that in the early stages of design it is important to explore extreme solutions, to think out of the box. The experts explained that ideas generated through brainstorming were more surprising. Although we thought that usefulness (workability, relevance) would be important for an innovative solution, it is not considered that important at this stage in the design phase. Therefore, the experts rated these ideas as more creative, despite they were not useful.

The issue of comparability of the design solutions needs further attention in future research. The judges commented that in the case of some design solutions (that were discussed after a prototyping session) they felt they missed relevant information. In the conversation the children pointed at objects and referred to options that did not require further explanation, at the time that information was redundant. In the transcripts however, the missing information hindered the judges somewhat in giving their opinion. Still we think that for this research the transcripts served their purpose. The alternative comparison of colorful but wobbly prototypes with childish writings would have given rise to bigger research issues. Nevertheless it is an issue that requires improvement.

Something that has not yet been addressed in this study and in the previous study [19] is the possible interference of intelligences required to empathize with the design case (the third level of required skills as described in our introduction of the framework.). For example, the interpersonal (social) intelligence may be required for empathizing with the incapacitated student. This intelligence is always required, independent of which method is applied. It could have inspired the children stronger than the intelligences required for the method. Furthermore it makes the relative difference in the number of required intelligences smaller.

**CONCLUSIONS**

Fifteen design master students assessed the creativity, novelty, non-obviousness, workability, relevance and thoroughness of ideas created by children in brainstorming and prototyping sessions. In contrast with what we expected, the design solutions generated in brainstorming sessions were judged as more creative than the design solutions from the prototyping sessions. Solutions from the brainstorming session were more novel and surprising; the solutions from the prototyping sessions were judged to be more relevant and workable. However, for all factors, apart from thoroughness, both the brainstorming and the prototyping ideas were on average judged on the positive side of the Likert scales, indicating that both methods lead to useful outcomes.

This reverse relationship between the number of intelligences involved in a design method and the creativity of the design solutions might be best explained by the process of triangulation: a larger number of intelligences strengthens the effect of triangulation of an increase in the amount of information in a smaller area of the design space and a decrease in the perceived creativity.

The results of this study and of our previous study in which outcomes were compared in terms of number of ideas and criteria are complementary. The previous study expressed the outcomes in terms of quantitative measures whilst the present study distinguishes between different creativity characteristics of design solutions.

The preliminary results from the second sessions indicate that the relative importance of the five criteria related to creativity depends on the design phase. In early design sessions, workability of a design solution is not yet required, novelty and surprise are more important before concept development.

Based on our findings we would advice design practitioners to use prototyping with nine-to-eleven-year olds instead of brainstorming. Prototyping provides sufficiently creative, yet relevant and workable design solutions, and at the same
time a more informative exploration of the design space in terms of options and criteria. The resulting ideas from both methods score positively on almost all criteria, and the relative difference between prototyping and brainstorming in terms of judged creativity is small.

Future studies, in which we will compare other design methods, are intended to extend our insights about the advantages and disadvantages of different early design methods. Our subsequent studies will focus on assessing what methods are optimal for different ages, depending on the development of the intelligences over time.

ACKNOWLEDGMENTS
Evan Karapanos, Karin Nieuwenhuizen and Janienke Sturm are thanked for their time and their advice.

REFERENCES