A Methodology to Evaluate Creative Design Methods: A Study with the BadIdeas Method

Paula Alexandra Silva
Lancaster University / Fraunhofer Portugal
Computing Department / Rua do Campo Alegre
LA1 4WA Lancaster UK / 4169-007 Porto, PT

Janet C. Read
School of Computing, Engineering and Physical Sciences - University of Central Lancashire
PR1 2HE, Preston, UK

ABSTRACT
The so-called creative design methods have become part of the everyday HCI-ers toolbox, however there is little discussion in the field concerning the actual value and the relative benefit of applying one method instead of another or of applying various methods in one's design efforts. These methods and techniques tend often to be applied in an unthoughtful uninformed manner.

This paper discusses the issue of evaluating and comparing the design methods and presents an overview of creativity measures for idea generation together with an attempt to rationalise those measures and combine them into a single value metric. This measure is then applied to assess the results obtained while using a specific method, the BadIdeas method, under various conditions; some observations and analysis on the possible effects of those conditions are performed.

Findings are surprising. Facilitated conditions positively affect participants' enjoyment of the method and the way they think about analysing products but the overall value of facilitation appears less than the overall value of unfacilitated work. The method seems to work better for groups that initially work individually, than those who start working in groups and overall results are better in a design, rather than in a redesign context.

AUTHOR KEYWORDS
User Experience Design, Creative Design Methods, Evaluation, Facilitator Effect, BadIdeas Method

ACM CLASSIFICATION KEYWORDS
H5.m. Information interfaces and presentation (e.g., HCI): Miscellaneous.

INTRODUCTION
It is common in HCI to engage users in design activities on the assumption that this participation will improve the uptake and acceptance of future technologies. In recent years, the users' involvement in design has expanded in such a way that users are frequently charged with generating not just the design solutions but also the ideas for it.

The answer on how to better support the development of better user experience design products is unclear. Philosophies such as User-Centred Design [23] and Participatory Design [1], which have the user's interests as a driving force [4] and which stress the importance of the active and creative participation of potential users in the design process [1] become increasingly relevant.

Yet, it is unclear, and there is a lack of empirical work on, how to drive present-day design activities, which methods work best in different contexts and how to evaluate aspects such as the effect of groups, the individual participants variability and the effect of facilitators. Moreover, assessing creativity and finding conditions that influence it is intrinsically hard methodologically.

This paper presents a study that aims at developing an understanding of these matters paying particular attention to the effect of aspects within a single method. The study involved 87 students using the BadIdeas method [3; 24], a tool for design that aims at nurturing creativity and innovation in the process of design and development of novel user interfaces. This paper reports the outcomes of this study; findings and conclusions are derived and discussed.

PREVIOUS STUDIES OF THE EFFECTIVENESS OF DIFFERENT EARLY DESIGN METHODS
The evaluation of design methods and of variations within design methods is complex and when researching the literature for methodologies to evaluate methods, these are scarce and often do little more than report on the difficulty of proposing one [5; 6]. Kelly [5] proposes a framework for comparing methods; nonetheless this framework mainly consists of a tabular synthesis of information with the main characteristics, advantages and disadvantages of the compared methods. Attempts to evaluate design methodologies are generally far from satisfactory, as there are too many variables, too much variety, and too little confidence in the results [6]. However, there are some measures that can be considered, such as the number of

OZCHI 2010, November 22-26, 2010, Brisbane, Australia.

OZCHI 2010 Proceedings ISBN: 978-1-4503-0502-0

ideas generated, the quality of the ideas generated, the engagement of participants, the learning of the facilitator, and the learning of the participants.

In many studies (e.g.: [7]; [8]; [9]), it is common to count the number of ideas generated, or to derive a ratio measure for ideas created when evaluating a method. This is a reasonable statistic for making comparisons of a single method across different instances of use and to determine how easy the method is to use across several groups of individuals. This is a quantitative measure, but this is not the norm in this domain. Christiaans [10] argues that "the enormous amount of research done in the last decennia highlights the lack of objective methods of assessment" of creativity in product design, a fact that Christiaans explains with the features of creativity, which are difficult to rate and therefore difficult to formalize into an objective instrument.

Several authors have proposed metrics to assess ideation effectiveness. Stammerjohan and Vance [11] propose three distinct but cumulative sets of measures: novelty, appropriateness and appeal. O'Quinn and Besemer' propose the Creative Product Semantic Scale [13] and Amabile's the Consensual Assessment Technique [14]. Christiaans [10] proposes seven metrics: creativity, prototypical value, attractiveness, interest, technical quality, expressiveness and integrating capacity and Shah and Vargas-Hernandez [12] propose four metrics: novelty, variety, quality and quantity. Despite the apparent difference of these various methods when analyzing the metrics each proposes, they are not so dissimilar. In most cases there is an aggregation of the concepts, that define creativity [15] [16]: novelty, which refers to unusualness and the capacity of causing one's surprise, and appropriateness, which refers to suitability and the capacity of distinguishing between the bizarre and the normal. For instance, in the Christiaans's set of metrics, creativity, attractiveness, interest and expressiveness would represent novelty, while prototypical value, technical quality and integrating capacity would represent appropriateness.

Another interesting measure is the engagement of participants, since this may contribute to the productivity of a group. Productivity is expected to increase with motivation, as the latter is expected to be directly correlated with engagement. Participants are engaged when an attitude of focus and commitment to an activity is observed. Csikszentmihalyi [17] defined this state as Flow.

Baruah and Paulus [9] studied the effects of training, on idea generation and concluded that training can increase both the quality and the quantity of ideas generated. Training here refers to the repeated exposure of people to a certain method.

Having verified the positive effect of facilitation, Putman [18] questioned if this effect was related to motivation or to efficiency. These two aspects have been studied by a number of authors with opinions divided. Putman [18] and Paulus and Brown [19] found evidence for the efficiency hypothesis. Kramer et al. [8] showed the positive effect of trained facilitators in face-to-face interaction, which was due to their motivational effects and to their ability to manage the interaction process effectively.

This paper tests the efficacy of several of the definitions identified above, proposes metrics that can be re-used to measure these different aspects and goes beyond this by proposing a single metric for the evaluation of these factors. Specifically, this single metric is used to investigate the BadIdeas method, presented in the following section, with the intention to further inform the use of the method. The choice of the BadIdeas method and its single use was due to the interest in understanding the way this method worked in itself.

THE BADIDEAS METHOD

The BadIdeas method [3; 24] is a tool for design that aims at nurturing creativity and innovation in the development of interactive solutions. In the BadIdeas method the transformation from a bad idea to a good idea encourages creativity and innovation.

There are several different ways of employing the BadIdeas method; determining which of these variations is most profitable in the design space is problematic.

The BadIdeas method was first presented in 2006 [2]. Since then, triggered by the analysis and findings of different instances of use, of a series of studies and across a number of contexts, the method has evolved across a couple of different versions in an iterative and incremental way [3]. Presently, the BadIdeas method is comprised of six phases (Figure 1), as described in each of the following paragraphs.

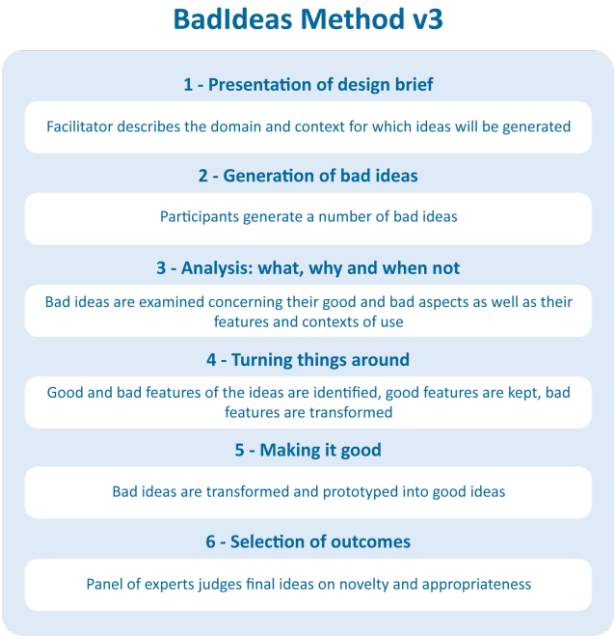

Figure 1: The BadIdeas Method v3 – Phases Synopsis

Phase 1 – Presentation of design brief – consists of the description of the design problem to the participants. This

includes a clear statement about the domain and the context of use for which participants will be generating ideas.

Phase 2 – Generation of bad ideas – consists of asking participants to generate bad ideas. This phase includes an explanation of what is meant by a bad idea and the provision of a couple of examples of bad ideas (for examples see [2]). The generation of a bad idea should be taken as a serious matter, and a bad idea intentionally aims at being bad, silly, crazy, weird and/or impossible [20].

Phase 3 – Analysis: what, why and when not – examines the nature of the ideas obtained in the second phase. This enables an understanding of these ideas as well as their related concepts and, ultimately, their design space. To support this process, the method uses a set of prompt questions (Figure 2).

THE BAD	THE GOOD
1 - What is bad about this idea?	1 - What is good about this idea?
2 - Why is this a bad thing?	2 - Why is this a good thing?
3 - Are there any other things that share this feature but are not bad?	3 - Anything that shares this feature but is not good?
4 - If so what is the difference?	4 - If so what is the difference?
5 - Is there a different context where this would be good?	5 - Is there a different context where this would be good?

Figure 2: BadIdeas Prompts

Phase 4 – Turning things around – uses a set of strategies to help participants to uncover new dimensions and possibilities for the bad ideas being examined. These strategies are: going back and forth, change the context and role-play. Going back and forth consists of alternating between positive and negative aspects of each feature. Change the context consists of looking at bad ideas from a different perspective and placed in different circumstances. Role-play of imagining that the bad idea was in fact deliberately designed as it is by an expert.

Phase 5 – Making it good – aims at turning the bad idea into a good one. This often happens on its own accord during phases 3 and 4, but if not, it is addressed as a separate stage. At this stage, participants must produce something that has the beginnings of pragmatics. This should be performed with the initial problem of the design brief in mind.

Phase 6 – Selection of outcomes – comprehends the evaluation and selection of the ideas that are going to be further developed into fully working prototypes. A panel of judges should evaluate the generated ideas and the low-fidelity prototypes, if available.

The study reported in this paper, and further detailed in the following section, observed the effect of the amount of facilitation, individual vs. group idea generation, instruction to participants, and context of ideation – Design or Redesign of the BadIdeas method. Part of a larger work on the BadIdeas method, this specific study was intended to discover how the manipulation of different aspects, such as the ones identified above, would affect the effectiveness and operation of the BadIdeas method.

STUDY DESIGN

The study reported in this paper was performed with undergraduate students from a suite of computer science related degrees studying at the University of Central Lancashire (Uclan), UK. Although the students were on different named courses, they were, at this stage in their studies, following courses in a common first year. 87 students were involved, 74 male, 13 female. Ages varied between 18 and 30 years old, with 37.9% being aged 18 and 32.2% aged 19.

Four main conditions – Design vs. Redesign, Facilitated vs. Not Facilitated, Start-up vs. No Start-up, and Group vs. Individual – were manipulated across the five studies in order to better understand the effects of the variables identified above. The result was that each of the five sessions had a slightly different shape:

- Design not facilitated (D)
- Design facilitated (DF)
- Design facilitated and with a start-up session (DFS)
- Redesign not facilitated (R)
- Redesign facilitated (RF)

The main author of this paper, the students, and the lecturer who would normally be teaching the students at the time of the BadIdeas method event attended each of the five sessions. The sessions were performed in two subsequent days in two similar rooms. All together, 26 groups were involved in the study: six in the D session, each with three participants; six in the DF session, one with two participants, one other with four participants, and four with three participants each; four in the DFS session, with four participants each; six in the R session, one with three participants and five with four participants each; and four in the RF session with three participants each. Table 1 synthesises this information and summarises the results for each group. The procedures and outcomes of these sessions are detailed next.

PROCEDURE

Each session began with the class lecturer introducing the researcher (the first author of this paper), to the class and outlining that the students would be taking part in this study. To guide the various sessions, the researcher followed a script (Figure 3). Students were informed about the activity they were going to participate in and the various stages in which the activity was going to be performed. Their challenge was to either:

- Design an interactive public display for the local town that could be used by children to stimulate healthy lifestyles, in the design (D) conditions, or

• Redesign the interface of an interactive application for a tablet PC to be used by children visiting museums, in the redesign (R) conditions.

Once the students got underway, they were either assisted or not according to the group they were in. Some groups in the class were instructed to work for the first five minutes individually and then in groups (denoted by an I in column 2 of Table 1), while others were instructed to work in a group from the beginning (denoted by a G in column 2 of Table 1). All the students were given paper materials to record their design ideas. Once the students had confirmed their final idea, they made a small videotape record describing it. At this stage they completed a short questionnaire that gathered their opinions of both participating in the session and their subjective evaluation of their own learning.

Students indicated whether or not they had enjoyed using the method on a Yes/No scale and if they would like to use the method again on a Yes/Maybe/No scale. To assess the impact the session had had on their thinking or learning about design, they gave a rating from Yes/Maybe/No in response to the question 'Do you think this method has changed the way you think about analysing products?'

Hi I'm palexa and we are going to do a design activity. We are going to design an interactive public display for Preston Train Station that could be used by children to stimulate healthy lifestyles//redesign the interface of an interactive application for a tablet pc to be used by children visiting museums

With the method we will use, we will have as many bad ideas as we can, then we will analyse what is bad (and good) about these ideas and we will then keep the good and change the bad to end up with a solution.

We will work in five stages:

Stage 1 – Fill in a small questionnaire about yourself (you can refuse to do this if you want).

Stage 2 – Generate bad ideas and write them down; then leave them on the table. Groups A, B and D use a single piece of paper to write all their ideas on, groups C, E, and F write on individual papers[1].

Stage 3 – Bring all the bad ideas of the group together – pick one for further development

Stage 4 – Make a short video that explains your final idea and tells us which of the original bad ideas you put into it…

Stage 5 – Complete the rest of the questionnaire

Figure 3: Script for the Study at Uclan, UK

ANALYSIS AND CODING

According to Kerr and Gagliardi [22], the Torrance Test of Creative Thinking (TTCT) [21] is the "best known standardized creativity test" and the one that "is supported by more evidence of validity". The TTCT assesses four creative abilities:

[1] Note that these instructions changed a little depending on the activity details.

• Fluency – the total number of interpretable, meaningful, and relevant ideas generated;

• Flexibility – the number of different categories;

• Originality – to the statistical rarity of the responses among the test subjects; and

• Elaboration – the amount of detail in the responses.

While recognizing the importance of the remaining metrics when analysing creativity, in this study, the authors only observed the results on fluency, i.e.: the number of ideas generated. This was the case, because the generation of ideas was not always performed the same way; some groups performed it individually, others in group. As a result, groups that started to work individually generated one list of ideas per participant and groups that worked in a group from the beginning generated one single list of ideas. This way, it was not possible to evaluate answers against flexibility, originality and elaboration, because ideas generated individually would have to be analysed in relation to the other individuals of the group, while the ones generated in group would need to be analysed in relation to the remaining groups. The analysis of this topic will be addressed in a different paper.

Three experts rated the final ideas on novelty and appropriateness. The group of experts comprised a person expert in HCI, a multimedia developer and an interaction designer who was employed in an industrial product design team. The three experts were not involved in the participatory sessions in any way, and each rated the final ideas individually on novelty and appropriateness by watching the videos the students had made and recording a numeric score using: Novelty – from 0 (not original) to 3 (never saw it or thought of it before) and Appropriateness – from 0 (not appropriate, impossible to implement) to 3 (possible to implement).

The subjective measures from the student questionnaire were coded as follows – each Yes response was coded 1, each No response was coded 0 and each Maybe response was coded 0.5.

EVALUATING DIFFERENCES

Wishing to go beyond a simple comparison of the numbers of ideas and the measured quality of the ideas, the authors sought to understand what measures were indicators of the overall value of the BadIdeas method in its various guises. In considering this problem, it was realised that metrics would be required to take into consideration all the dimensions which are affected as a result of changing conditions, such as whether the group was facilitated or not.

Four metrics were selected and defined to compare the different dimensions. These were:

• Mass of Ideas generated (MI) – the number of ideas generated by a group / the number of participants

The Mass of Ideas generated (MI) is the number of ideas generated, normalised by the size of the group. The size of the group is often a point of interest in design sessions, and

often researchers wish to compare how this affects the outcome. By using this metric, it is possible to compare outputs from groups of different sizes. MI has the potential to be very large but, given the time constraints of an average design session, it is expected it will not normally go into double figures.

• Quality of Ideas generated (QI) – the average novelty rating (from the three evaluators) * the average appropriateness rating (from the same three evaluators)

The Quality of Ideas generated (QI) is a product of the idea's novelty and appropriateness ratings. An idea can be unique and groundbreaking, but at the same time wholly inappropriate to the audience or the environment. By multiplying the two values, we can indicate whether the idea is novel and appropriate in a single number, in an increasingly significant manner. Note: with the ratings used in this study, the range of values for QI is from 0 to 9. It has to be stressed that as QI is then not interval scaled and lends an advantage to ideas that are both highly novel and highly appropriate. This is intentional as it is the quality of ideas, rather than the number of ideas that is key to innovative design. It is acknowledged that the importance of novelty versus appropriateness may depend on context: but in the design problems outlined in the studies reported in this paper they are considered to have equal weight.

• Enjoyment of Participants (EP) - recorded as an average of the scores for whether the student would like to do the activity again and their self rating of enjoyment (both scored from 0 to 1)

The Enjoyment of Participants gives a value between 0 and 1 for the reported student experience. Earlier work by one of the authors [3] has shown that a single rating of enjoyment is often flimsy and so the product of these two ratings, both using different angles, is a more robust measure.

• Learning of Participants (LP) – recorded as a single rating – based on the ratings given by students about whether or not carrying out the activity had changed their ideas about design as a process and activity (scored from 0 to 1)

The Learning of Participants results in a value between 0 and 1 for the reported student learning. It is highly possible that this will not be an especially robust metric. A better measure might be gathered if students were to have a second turn at the method and in that case it might be related to a positive change in the quantity and quality of ideas from one session to the second. It is noteworthy that in the groups that had a start-up session, i.e. a practice run, the average number of ideas was lower than for the sessions without a start-up session. This aspect would need further investigation to identify if this is expected in all instances of application of the method.

For this study, as a single facilitator attended all sessions, the Learning of the Facilitator (LF) was not included but had it been included in the single value metric it would have been counted in the same manner as Learning of Participants (LP), with a questionnaire.

THE VALUE METRIC (VM)

Combining the five individual ratings (in this study only four were used), a single calculable value metric that can be applied to any design session. This metric is as follows:

Value of Method (VM) = MI + QI + mEP + q(LP + LF)

In this formula, m and q are constants; m is taken as the average value for the mass of ideas (MI) generated across the groups per method. As a result of this study an initial constant for the general use of the BadIdeas method (m) is derived. q is similarly the average value for the quality of ideas (QI) generated across the groups for this method. These constants are used to attribute some weighting to the enjoyment and learning dependent on the specific design method.

It is assumed that there is some correlation between the number of ideas produced and the enjoyment, i.e.: although a participant will still report in a 0 - 1 scale his or her enjoyment, the general value of this might be higher in a productive session. Similarly, the constant for learning is factored by the average quality rating for the activity as – despite the students' own view - there will generally be more learning where the quality of outputs is high. By calculating these constants for each method, it should be possible to compare the value of different methods. In this study, the constants for the BadIdeas methods m and q were 2.88 and 2.06 respectively. As it is the first time such constants have been calculated for the BadIdeas method, it is acknowledged that these values will require refinement with further studies.

RESULTS

The following table summarises the results for each group.

Group	Detail	MI	EP	QI	LP	VM
Z	A3GD	4.00	0.25	2.78	0.00	7.50
Y	A3GD	0.67	0.67	4.67	0.50	5.33
X	A3GD	1.33	0.75	5.33	0.67	8.21
W	A3ID	4.33	0.25	2.67	0.67	7.00
V	A3ID	5.33	0.67	7.11	1.00	12.44
U	A3ID	4.33	0.17	7.11	0.33	11.44
T	A4GDF	1.75	0.50	0.00	0.63	1.75
S	A3IDF	1.33	0.58	2.67	0.00	4.00
R	A3GDF	2.00	0.75	2.22	1.00	4.22
Q	A3GDF	2.00	0.83	2.33	0.83	4.33
P	A2IDF	5.50	0.75	0.00	0.50	5.50
O	A3IDF	4.33	0.75	1.56	0.50	5.89
N	A4GDFS	2.25	0.81	2.22	0.50	4.47
M	A4GDFS	0.50	0.75	0.67	0.63	1.17
L	A4IDFS	3.00	0.56	0.67	0.63	3.67
K	A4IDFS	1.25	0.63	2.78	0.50	4.03

J	B4GR	1.75	0.94	0.44	0.63	2.19
I	B4GR	0.75	0.06	0.00	0.25	0.75
H	B4GR	1.00	0.25	1.78	0.00	2.78
G	B3IR	7.00	0.00	1.56	0.00	8.56
F	B4IR	2.25	0.75	1.00	1.00	3.25
E	B4IR	3.00	0.19	2.67	0.38	5.67
D	B3GRF	2.33	0.00	0.44	0.00	2.78
C	B4IRF	2.75	0.75	0.00	0.50	2.75
B	B3IRF	6.67	0.58	0.00	1.00	6.67
A	B3IRF	3.33	0.50	1.00	0.50	4.33
	AVG	**2.88**	**0.53**	**2.06**	**0.50**	**5.03**
	SD	1.83	0.28	2.04	0.32	2.89

Legend: A/B – Design brief challenge; Digit (e.g. 3) – Number of group participants; I – Individually at the beginning; G – Group; D – Design not Facilitated; DF – Design Facilitated; DFS – Design Facilitated with Start-up Session; R – Redesign not Facilitated; RF – Redesign Facilitated

Table 1: Summary of Group Details and Results

Figure 4 illustrates the data from Table 1. It can be seen that there was considerable variation across the groups but it can also be seen how the Value Metric (VM) follows closely the general trends for MI and QI whilst at the same time accentuating the differences.

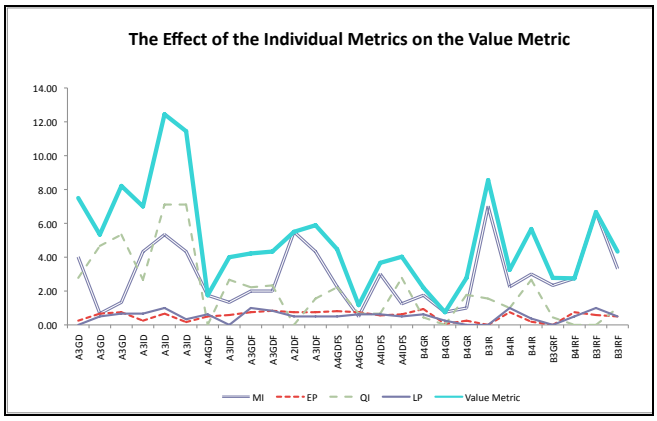

Figure 4: The Effect of the Individual Metrics on the VM

When each condition is considered separately it is clear that there are some differences. As shown in Figure 5, on average the conditions design, not facilitated, individual and no start up – resulted in better average marks than the others. This is not conclusive as the samples considered were too small, especially in the case of start up vs. no start up, and because of this, there are no claims made here for statistically significant differences, but in the absence of evaluative data in most currently published design studies, these possible differences are a starting point for discussion.

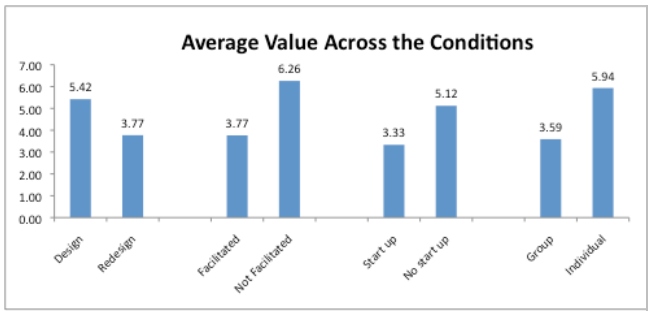

Figure 5: Average Value Across Conditions

Two aspects are worth mentioning, one regarding facilitation and another regarding the quantity of ideas. Concerning the former, and as Figure 6 shows, participants' enjoyment of the method is higher in facilitated conditions. This provides some evidence to Putman's [18] hypothesis that facilitation may be related to motivation. Therefore, even if in the Not Facilitated condition, better marks are obtained, facilitation should not be assumed to be a poor addition to a design session.

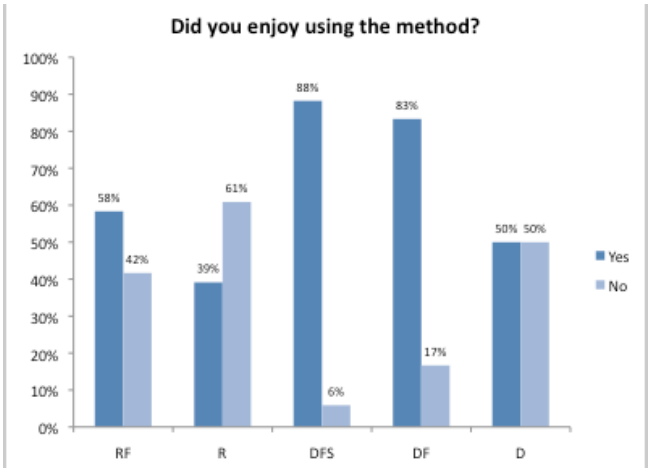

Figure 6: Effect of Facilitation in Enjoyment of the Method

Concerning the quantity of ideas, it is important to note that even if participants generate more ideas while performing idea generation individually, these may eventually overlap with the ones of other participants in the same group – this can present itself in this metric (and indeed in other metrics) as an inflation of the individual effect on idea creation (MI). It would be necessary to take out all similar ideas and recalculate the number of individual ideas to further investigate these phenomena. Besides, while working in groups, participants are more subjected to the social effects of creativity, such as evaluation apprehension, production blocking and free riding.

DISCUSSION

The discussion here is in two sections. The first section is a reflective consideration of the method used in this study to derive empirical data relating to the different aspects of creative design sessions. The second section goes beyond the quantitative metric, by considering the slice in time of the BadIdeas session where a specific bad idea is proposed,

and discusses the evaluation of BadIdeas from the point of view of effectiveness of the bad idea for use.

The Value of Method Formula

The formula that is presented in this paper is a first attempt to quantify design sessions and to allow some comparison across different conditions of use. In this study, the value metric appears to have assisted in the differentiation of contexts and this has been shown to be useful. It is not possible, in the scope of this paper, to comment on how the value metric fares with other methods but an investigation of this is ongoing.

When different methods are being compared, for example BadIdeas and Brainstorming, there will be a need to perhaps explore the expected outcomes while reflecting on the value. For instance, in BadIdeas the intention is to generate bad ideas, which is not always easy [3; 24], in brainstorming the intention is to begin with many ideas, so there might be a high average value for MI. Ideas are then hone down, generally to quite a functional idea. The functionality of this idea might make the achievement of high values of QI difficult. It might be the case that these factors will cancel one another out in a comparison using the method constants m and q. Further investigation is required to establish firstly, a robust constant for the BadIdeas method, and secondly, whether similar constants can be calculated for other methods.

Evaluating Ideas Generated

It is complex to identify and to describe what constitutes a good bad idea. When reflecting upon the examples of bad ideas generated during a number of BadIdeas sessions, some aspects emerge. These aspects include the clarity of the bad idea statement and the potential of transformation of the bad idea chosen to proceed with.

Concerning the clarity of the bad idea statement, the idea, which is in the head of its creator, is different from the idea that its creator externalizes and is finally communicated. Moreover, it normally needs to be written down. If the externalization of a bad idea is not straightforward to its own creator, it is even harder for an external person to fully understand what is in the head of the idea creator. Because of this, the process of analysing an idea is challenging, as there is a breakdown between the idea that is in the mind of its creator and the idea verbalized or understood by a third person.

As to the potential of transformation of an idea, this is a crucial attribute of a good bad idea. Transformation is in the basis of the BadIdeas method. This process is closely related to phase 3 – Analysis: what, why and when not – of the BadIdeas method, as transformation greatly depends on the understanding of a given bad idea. When developing an understanding of the bad idea, participants have to identify: what good bits can be kept, what bad bits can be transformed, what neutral things can be changed – and then the process is to identify what is good and keep it and then identify what is bad and change it, i.e. transform it into something good given a certain design brief.

There are some characteristics and attributes that a bad idea should include:

• A good bad idea has to be purposely bad, silly, crazy, weird and/or impossible

• A good bad idea has to be vague enough to allow transformation

• A good bad idea is not too detailed, that the creator finds it hard to lay off of those details

• A bad idea does not need to be related to anything or to any domain in particular, including the design brief

The effect of the 'badness' of an idea on the quality of the eventual idea selected is one area that is really interesting for the field of User Interface Design. It is traditional to look at design sessions as a single box process and the impact of the midpoint is often lost. This is an area that is currently being further investigated by the authors of this paper.

CONCLUSION

HCI comprises and makes use of a vast palette of methods; however HCI is not strong when it comes to the evaluation of those methods and specifically the ones for creativity and innovation. However evaluation and comparison is needed not only when one is required to select a specific method to work with but also when one wants to compare the results of different methods. Moreover, when one is referring to early design methods, creativity comes into play; this adds complexity to the initial problem, since both the assessment of creativity and the study of how different conditions affect it are methodologically hard.

To 'quantify' 'better' in this context, of evaluating and comparing across methods, a metric has been derived that goes beyond simply counting the number of ideas and their perceived quality, and includes some factors relating to the enjoyment of the users and the learning across the activity. Specifically, the metric includes the number of ideas generated, the quality of the ideas generated, the engagement of participants, the learning of the facilitator and the learning of participants are considered. The metric proposed was then applied to the BadIdeas method.

This paper also presented an overview of the BadIdeas method as it is currently used for design ideation. As well as presenting the method in a practical way, this paper has also unpicked the idea of a bad idea by exploring what it is that makes a good bad idea. The BadIdeas method has been evaluated in a study with deliberate manipulation of some of the aspects of design that are commonly varied from study to study. Results demonstrated that there is some variation and indicates that certain 'types' of design sessions might culminate in better results.

Having only been used in a single method, indeed, in a single instance of the method, to compare the manipulation of settings, the value metric will need further use and further refinement. Initially it will, if it is to be applied across instances of use, be necessary to investigate the

robustness of the constants and to consider how constants for a single method should best be derived. Taking the metric to new methods will be a second stage.

In the particular situation of the BadIdeas method, this paper has shown that, with the design and redesign challenges of this instance, the BadIdeas method performs better in a design context than in redesign, with individual idea generation at the first stage, and facilitation seems to have no overall effect on the ideas but does seem to make the sessions more enjoyable.

There are many aspects of this study that remain open for future investigation. The creativity the student brings into the group session is currently being studied and it is anticipated that this creativity brought in might have an effect in certain conditions. Further work might investigate the impact of the specific design or redesign idea on the ideas generated and the collaborative effect of group work on the ideas.

ACKNOWLEDGMENTS

Thanks go to all the participating students and to the students and staff from University of Central Lancashire.

REFERENCES

[1] Kyng, M. (1994). Scandinavian Design: Users in Product Development. Proceedings of CHI '94, Boston, USA.

[2] Dix, A., T. Ormerod, M. Twidale, C. Sas, P. A. G. d. Silva and L. McKnight (2006). Why Bad Ideas Are a Good Idea. HCIEd2006, Limerick, Ireland.

[3] Silva, P. A. 2009. Designing User Interfaces with the BadIdeas Method: Towards Creativity and Innovation. Department of Computer Science. Lancaster, Lancaster University. Doctor in Philosophy.

[4] Preece, J., Y. Rogers and H. Sharp (2002). Interaction Design: Beyond Human-Computer Interaction. USA, Wiley.

[5] Kelly, J. C. (1987). A Comparison of four Design Methods for Real-Time Systems. International Conference on Software Engineering Monterey, California, United States, IEEE Computer Society Press.

[6] Kelly, S. R., E. Mazzone, M. Horton and J. C. Read (2006). Bluebells: A Design Method for Child-Centred Product Development. NordiCHI 2006: Changing Roles, Oslo, Norway, ACM.

[7] Oxley, N. and M. Dzindolet (1996). "The Effects of Facilitators on the Performance of Brainstorming Groups." Journal of Social Behavior and Personality 11(4).

[8] Kramer, T., G. Fleming and S. Mannis (2001). "Improving Face-to-Face Brainstorming through Modeling and Facilitation." Small Group Research 32(5).

[9] Baruah, J. and P. B. Paulus (2008). "Effects of Training on Idea Generation in Groups." Small Group Research 39(5): 523-541

[10] Christiaans, H. H. C. M. (2002). "Creativity as a Design Criterion." Creativity Research Journal 14(1): 41-54.

[11] Stammerjohan, C. A. and D. Vance (2001). New Product Development: A Comparison of Traditional Creative Methods with Genetic Algorithms in the Ideation Stage. Proceedings of the Society for Marketing Advances 2001 Conference, New Orleans.

[12] Shah, J. J. and N. Vargas-Hernandez (2003). "Metrics for Measuring Ideation Effectiveness." Design Studies 24(2): 111-134.

[13] O'Quin, K. and S. P. Besemer (1989). "The development, reliability and validity of the revised Creative Product Semantic Scale." Creativity Research Journal(2): 267-278.

[14] Amabile, T. M. (1982). "Social psychology of creativity: A consensual assessment technique." Journal of Personality and Social Psychology(43): 997-1013.

[15] Perkins, D. N. (1988). Creativity and the Quest for Mechanism, Cambridge University Press.

[16] Martindale, C. (1999). Biological Bases of Creativity. Handbook of Creativity. R. J. Sternberg. Cambridge, Cambridge University Press: 137-152.

[17] Csikszentmihalyi, M. (1996). Creativity: Flow and the Psychology of Discovery and Invention. USA, 1st HarperPerennial.

[18] Putman, V. (2001). Effects of additional rules and dominance on brainstorming and decision making. Arlington, University of Texas at Arlington. Ph.D: 6026.

[19] Paulus, P. B. and V. R. Brown (2003). Enhancing Ideational Creativity in Groups: Lessons from Research on Barinstorming. Group Creativity: Innovation Through Collaboration. P. B. Paulus and B. A. Nijstad. USA, Oxford University Press: 110-136.

[20] Dix, A. (2000). "Silly Ideas." from http://www.comp.lancs.ac.uk/computing/users/dixa/resmeth/away-day/bad-ideas.html.

[21] Torrance, E. P. (1966). "The Torrance Tests of Creative Thinking: Norms-Technical Manual." Personnel Press.

[22] Kerr, B. and C. Gagliardi (2002). Measuring Creativity in Research and Practice, Arizona State University.

[23] Norman, D. A. and Draper, S. W. (1986) User Centered System Design; New Perspectives on Human-Computer Interaction, L. Erlbaum Associates Inc.

[24] Silva, P. A. (2010) BadIdeas 3.0: A Method for Creativity and Innovation in Design. DESIRE 2010: First Conference on Creativity and Innovation in Design, August, Aarhus, Denmark.