Quantitative Evaluation of the Effectiveness of Idea Generation in the Wild

Lassi A. Liikkanen[1,2], Matti M. Hämäläinen[2,3], Anders Häggman[4], Tua Björklund[2], and Mikko P. Koskinen[2,3]

[1] Helsinki Institute for Information Technology HIIT, P.O. Box 19215, FI-00076 AALTO, Finland, Aalto University and University of Helsinki
[2] Aalto Design Factory, P.O. Box 17700, FI-00076 Aalto University, Finland
[3] Engineering Design and Production, P.O. Box 14100, FI-00076 Aalto University, Finland
[4] Dept. of Mechanical Engineering, Massachusetts Institute of Technology, 77 Massachusetts Ave, 3-449B, Cambridge, MA 02139, USA
{Lassi.Liikkanen,Matti.M.Hamalainen,Tua.Bjorklund, Mikko.P.Koskinen}@Aalto.fi, haggman@mit.edu

Abstract. New ideas are the primary building blocks in attempts to produce novel interactive technology. Numerous idea generation methods such as Brainstorming have been introduced to support this process, but there is mixed evidence regarding their effectiveness. In this paper we describe an experimental, quantitative methodology from the domain of product design research for evaluating different idea generation methods. We present prominent results from relevant literature and new data from a study of idea generation in the wild. The study focused on the effects of the physical environment, or in other words, the physical context, on designers' capacity to produce ideas. 25 students working in small groups took part in an experiment with two design tasks. Moving from an office environment to the actual surroundings of the intended use, we discovered that the change in resulting ideas was surprisingly small. Of the measured dimensions, the real-world context influenced only the feasibility of ideas, leaving quantity, novelty, utility and level of detail unaffected. This finding questions the value of diving into the context as a design idea generation practice.

Keywords: Design methods, idea generation, creativity, psychology.

1 Introduction

Finding the right direction, that is, defining what to design is the evident challenge in most development projects. This also holds for the design of new interactive technologies. In attempts to organize creative efforts effectively and maximize their impact, many research and development (R&D) organizations employ innovation management. A variety of methods have been introduced to ensure quantitatively and qualitatively sufficient results during the fuzzy front-end of design. These methods, called idea generation (IG) techniques, and comparable methods for requirements elicitation have been available for some time. In the tradition of this craft, formal IG methods such as synectics, six thinking hats, morphological analysis, lateral thinking,

M. Kurosu (Ed.): Human Centered Design, HCII 2011, LNCS 6776, pp. 120–129, 2011.

TRIZ, and naturally Brainstorming [1], can be found. The variety of available methods is nowadays considerable if we also include a category called creativity support tools [2] to the collection of ideation techniques.

Despite the extent of idea generation methodology, we and many others find something essential missing from the debate. The presentation of new ideation methods commonly lacks an objective assessment of the impact that the tools and practices are supposed to yield. They lack proof regarding their effectiveness and suitability for different types of projects. This has led design researchers and psychologists to question the effectiveness of these procedures. Indeed, Brainstrorming alone has received attention from psychologists since its introduction in the 1950’s [see, for example, 3-5]. There are many approaches to proving the effectiveness of ideation. One could target social interaction and communication [6], provide self-evaluation instruments for the designers [7], or analyze the output of the process as will be done in this study. The historical source of inspiration for the latter approach is the creativity measurement literature [8] from which most of the methods are derived from.

In this paper, we discuss the quantitative *idea analysis methods*. The aim of this methodology is to focus on the qualities and quantities of ideas produced in experimental settings. With this approach, we can investigate how different controlled and manipulated variables influence the outcomes of the IG process. As an example of applying this methodology, we include an experiment studying the effects of design surroundings in generating new interactive product-service concepts.

2 Studying Idea Generation through Produced Ideas

Idea analysis has so far been utilized within different fields of psychology, namely cognitive [9] and social [10], but also in management science [11]. These studies have dealt with IG as a generic topic, assessing ideas for large-scale social and environmental problems. Nowadays there exists a wealth of studies that have particularly addressed IG in the context of design. They most often concern product design and have been carried out in the engineering domain [12-14]. However, they are relevant for interaction design as well, given that interaction design is a design discipline too [15]. For instance, the research findings repeatedly challenged the notion of group work’s efficiency in idea production in comparison to nominal groups consisting of independent individuals. These findings on the illusion of group productivity are paralleled by the discovery of design fixation, a phenomenon caused by exposure to inspirational material.

For successful empirical research it is important to have a theoretical basis to help make sense of the experimental data. The majority of studies share a psychological perspective in which IG is considered a predominantly memory-based activity. There are several theoretical models on how new ideas can be produced based on re-using knowledge that designers have acquired previously [10, 16, 17]. These models describe ideation as knowledge reuse and refer to psychological concepts such as memory recall and conceptual combination. With these models, we have just started to understand the dynamics of idea generation as it occurs in reality. This means abandoning the romantic or mystic accounts of ideas just emerging in one’s mind.

2.1 Research in Practice

Design IG studies that apply idea analysis methods [see, for instance 13, 18] are typically experimental psychological studies. The goal is to determine causal relations; i.e. what consequences do changes in the fixed factors (*independent variables*) have in terms of certain outcomes (*dependent variables*). The effects should be explained by referring to the properties of the background theory, i.e. human cognition. IG studies might be more accurately described as quasi-experimental research. This refers to the fact that the assignment of individuals to groups may not be truly random, and to external variables that cannot be controlled which may affect the data collection. A major reason for this is that the participating designers are never completely identical to each other in terms of all the relevant factors in their background experience.

The idea analysis method has a few requirements. First of all, there must be a lot of data. An interesting repeated finding in design IG research [17], as well as from creativity tests [8], is that independent minds will repeatedly generate similar ideas. The empirical evidence is commonly acquired in one or more successive design sessions. In each session, a design brief advices the participants to generate ideas for a given situation. In design studies, a visual representation of the idea is usually preferred. The sessions are commonly managed and facilitated by the experimenter(s). An example of an experimental IG study setup would be to investigate how the group size (independent variable) affects the amount of ideas produced by one person (dependent variable). This has been the prevalent choice in studies of Brainstorming, which have often compared real, interactive groups with nominal groups, equal in size, but consisting of individuals working in isolation [see, for example, 4, 5]

Experimental research on design IG also requires limiting the freedom of designers in order to maintain control over the experiment. The participants need to follow quite strict rules and adapt to working in ways new to them. This can be problematic due to the fact that it induces some learning effects. There are also some concerns regarding subjects who have been traditionally selected by convenience. The majority of published design studies compare groups of (design) students rather than professionals. This is problematic to some extent, as there are known effects of gaining skills and knowledge, i.e. expertise in design [19]. However, for IG research, a uniform education and lack of design experience may be beneficial if it results in greater homogeneity of background knowledge from which the ideas emerge, and consequentially create more pronounced experimental effects.

After the design experiment is carried out, the analysis proceeds by measuring the dependent variables. The analysis always requires some domain expertise to understand and interpret the ideas. This is usually carried out by select domain experts, but recently interesting proposals about crowd-sourcing have also been suggested [20]. Typical dependent variables are *Quantity*, *Creativity*, *Feasibility*, *Elaboration*, and *Categorization.* The most commonly used variable is the number of ideas produced, labeled quantity or productivity. Typically only unique, non-repetitive ideas are counted for this measure. This simple figure is often complemented by some measure of creativity. Although challenging to define [21], creativity is usually assessed through *novelty* and *quality* of the idea. Novelty, diversity, or commonality all represent how common the idea is. This can be referenced either to the pool of ideas collected or to solutions known to a domain expert (the evaluator). Quality can be operationalized as functionality, feasibility, or even usability. Feasibility and elaboration are quite straightforward variables describing the readiness for implementation and the detail of description. Elaboration is generally negatively associated with quantity. Finally, categorization refers to assigning ideas to clusters by similarity. This is theoretically justified and provides a measure of novelty [for more information, see 12, 14, 22].

An important question regarding the dependent variables is that are they valid? Do they measure things that we think are important regarding the success of ideation? This is a difficult question because the application of any of these measures is a compromise. Ultimately, we are interested in the ideas that can lead to breakthrough innovations. But that is nearly impossible to measure. Thus far, many researchers have adopted the thinking that quantity breeds quality – having many ideas guarantees finding a few good ones [23]. Some researchers believe that only the number of "good" ideas is indicative of how well IG succeeds [24]. Our present approach is pluralistic, we analyze all unique ideas in great detail.

The idea assessment process requires reliability and robustness. Thus several independent evaluators are needed, and their agreement should be confirmed using inter-rater agreement assessment (kappa statistic) or consensus. The evaluation process can be iterated a few times particularly if the problem area is new and the ideas are not predicted. However, ultimately the space of potential solutions for the given task becomes mapped extensively, and the assessment of individual ideas is easier as they tend to fall into predefined categories.

3 Essential Findings from Idea Analysis Research

3.1 The Dark Side of Brainstorming

After a long era of investigations into the nature of Brainstorming, its positive and negative aspects have become well-known. They point out that while properly organized Brainstorming does tease out a huge amount of creative potential, there are many kinds of hindrances as well. Group situations generally induce phenomena known as *production blocking*, *free riding*, *evaluation apprehension* [3], and different forms of *group think*. Production blocking refers to the fact that once someone else is taking a turn to express an idea, others are expected to listen. This blocks both their output and, if truly paying attention to the presenter, their idea production, as well. The bigger the group, the bigger the loss due to production blocking. It has been repeatedly shown that real groups of more than two people always underperform in comparison to nominal groups [3]. This has motivated the development of different electronic brainstorming platforms, which overcome the production blocking issue and allow multiple people to contribute simultaneously [25].

Social loafing or free riding is a common feature of unequal distribution of effort induced in a group. Participants may decrease their efforts if they think they do not have anything to add, or if they believe others will take care of it. The fear of negative feedback is labeled evaluation apprehension, and mainly effects the communication of ideas. Putting out an idea requires good self esteem or high confidence in the group treating it with an open mind. These three factors directly influence the quantity of work a group can produce. Finally, there are other forms of group think, which affect the qualitative nature of the produced ideas. The best example is the influence of a "group leader" (whether by appointment, charisma, appearance, or social ranking), who can dominate the group opinion. [26]

3.2 Individual Level Effects

Changes in the IG process are commonly described with two distinct types of effects: stimulation and fixation. *Stimulation* is the positive effect that having outside influences prior to or during IG has. Many design practitioners share the view that all sorts of external material can inspire the production of new designs. With some precautions, this seems to be the case [27], although the evidence from design is still scarce [13]. The stimulation is believed to help designers to get started with the ideation process. It helps mainly to create ideas that are similar to the stimuli. For this reason, heterogeneous stimuli are preferred, and have been found to facilitate the production of more diverse ideas.

Fixation, on the other hand, is the down side of stimulation. It refers to the negative impact of external influence. A number of studies under the label of "design fixation" have tried to characterize how fixation shifts the design IG process to a certain, possibly unwanted, direction [28, 29]. They define design fixation as a tendency to unknowingly reproduce parts of the given examples to a new design. This occurs even if the examples contradict the design requirements. In the psychological literature [9, 30] this is called unconscious plagiarism.

IG is an activity that naturally evolves over *time*. A recent review identified three kinds of effects that time can have on idea generation processes [31]. Time pressure, total duration, and time decomposition were each considered important for the effectiveness of an ideation session. Time pressure has a non-linear relationship to productivity and creativity, and the reviewed studies suggest that a hectic warm-up session can affect the productivity of a later, more relaxed real session positively, in terms of productivity and creativity. Interestingly, productivity over time shows an initial burst of productivity when the most common ideas are produced. Time decomposition refers to splitting the ideation session to cope with the time-related issues. For instance, by having a break or by dividing the session between individual and interactive work phases can help.

4 An Investigation of Ideation in the Wild

4.1 Background

In the spirit of user-centered development, a straight line between the world of users and developers is desired. An open question with IG practices in the professional setting is the evident distance between the site of use and design. As a solution, some methods of facilitating the designers' stepping into the users' world have been described. For conceptual design, IDEO has introduced the Deep Dive method, and Oulasvirta et al. have described a Bodystorming ideation method for mobile context [32].

We wanted to investigate the effects of creating new software service concepts in real environments. Using two locations and two tasks, we explored how the change in physical context influenced the resulting design. In the experiment, we had one central research question in mind: Are there systematic differences between the contexts regardless of the ideation task? The cognitive theories of IG would generally make two contradicting assumptions: the stimulation potential of the real environment should qualitatively change the ideas, but it influences designers' thinking to the existing solutions and the specifics of the environment [10, 18].

4.2 Method

Participants and experimental design. We recruited 25 university students participating in a cross-disciplinary product development intensive course (14 male, 11 female). Their average age was 25.9 (st. dev. 4.8 years) and they all represented different disciplinary backgrounds, coming from 14 different nationalities. All subjects evaluated their English skill as fluent or average. After their written consent, they were split into groups of three people. The division was arranged so that the heterogeneity of the groups was intentionally maximized by controlling the mix of backgrounds (design experience), sexes and nationalities.

We examined the effect of context on IG using a factor called *Physical location* trialed across two tasks. For the tasks, we had a Hotel and a Gym task. The former dealt with new services for lobby touch screen points and the latter with an "intelligent gym" in general. Both of the tasks were either carried out at the typical *office* location or at the associated *real-world* site. Participants were not informed about the nature of this experiment, other than that it was a competitive IG practice. Every group worked once in the real setting and once in the office, with a different task at each location. The order of the tasks was counter-balanced between the groups and distributed pseudorandomly.

Tasks and Materials. The participants were asked to create ideas at two consecutive locations. On average, they had a 60 min break between the tasks allocated for transportation (all locations were located within a 500 meter radius). The IG was divided between an individual IG (5 min) and a group IG phase (15 min). Before the IG phases the participants had 5 minutes to explore the context in the real-world setting (see Figure 1) or inspect photos of the real context in the office setting.

Fig. 1. Inspirational photos of the intended use context. The same locations were explored by the participants in the real context situation.

Fig. 2. Participants working in the gym setting (left) or in the office setting (right) in the group brainstorming phase

4.3 Results and Analysis

Here, we only present the results from the individual working phase, in order to avoid potential effects that the type of context had on group social interaction which might obscure the context effect on the produced ideas themselves. We analyzed all qualified ideas (N=139) produced by the individuals according to the dimensions of novelty (NOV), utility (UTI), feasibility (FEA), level of detail (LOD), mode of presentation (MOP) and categories. The assessment scales were defined bottom-up by examining the data on two iterations, so that five evaluators first evaluated a subset of all ideas for both tasks. After this they discussed the suitability of the proposed measures for the data and then converged on the scales. Some unidentifiable ideas were discarded. For all dimensions except MOP, a four step ordered scale used (values 0-3). On the following round, two independent evaluators used the revised scales to score each dimension of the data. The hand-coded data was imported into SPSS statistics software and analyzed using non-parametric statistical tests. Samples of the produced ideas are presented in Figure 3.

Fig. 3. Three examples of documented ideas

We first examined the Gym and the Hotel lobby tasks for between groups effects. Interestingly, no statistically significant differences were discovered in the quantity of the ideas or in the evaluated dimensions. On the individual level, we tested whether subjects differed between the contexts across the two tasks. We found few changes across conditions, the only statistically significant difference occurred in the feasibility dimension, which was generally higher in concepts produced in the real-world setting (see Table 1 below). No changes in the categorical distribution for the mode of presentation were found across the conditions.

Table 1. Results from a Kruskal-Wallis test assessing the differences across all produced ideas in both experiments

	Novelty	Utility	Feasibility	LODetail
Chi-Square	.255	.088	5.332	.918
P-value	.613	.767	**.021**	.338

5 Discussion and Conclusion

In this paper we have presented the paradigm of idea analysis in design IG for the interaction research community. Having first discussed the earlier results found in literature, we described our own experiment on idea generation in the wild. In our study, we witnessed surprisingly small differences in the quantity and quality of ideas produced in different physical contexts. However, we noticed that the feasibility of ideas tended to increase if they were produced in the real use context. This is best understood through the constraining effects of environment. Even though we expected to see positive effects from stimulation, the only observed consequence of the location was getting increased grounding to the present environment. The outcome supports detachment of the target environment as an evidence-based practice in design idea generation during concept development. This is compatible with a theory of idea generation in which all additional pieces of information around the design brief accumulate mentally to design requirements. These design constraints can have a negative effect on a process such as conceptual design which should aim at maximum creativity of the output.

In the future, we plan to make a thorough qualitative analysis of the additional group work data and video material, which were neglected in this study in order to focus on the quantitative results. We see plenty of room and need for similar, theoretically motivated investigations about HCI design methodology in the future. The present investigation had limited statistical power and leaves room for future replications around the topic. The methodology documented in the paper provides for easy follow ups and instructs the reader in quantitative assessment of idea generation methods. We hope that the present study will spark the creation of new approaches on how to leverage exposure to real context environments, to the advantage of designers, maybe through means other than brainstorming *in situ*. However, our study clearly shows that qualitative and quantitative gains cannot be simply assumed to follow.

Acknowledgments. We thank the participants, as well as Juha Forsblom, Viljami Lyytikäinen and Jussi Hannula at the Aalto University Design Factory for helping with the arrangements. The preparation and presentation of this paper was supported by *LUTUS – Creative Practices in Product Design* - research project. The effort of co-authors in the preparation of this paper was equally shared.

References

1. Osborn, A.F.: Applied imagination: principles and procedures of creative problem-solving. Revised edition. Scribner, New York (1957)
2. Shneiderman, B.: Creativity Support Tools. Communications of the ACM 50, 20–32 (2007)

3. Diehl, M., Stroebe, W.: Productivity loss in brainstorming groups: Toward the solution of a riddle. J. Pers. Soc. Psychol. 53, 497–509 (1987)
4. Diehl, M., Stroebe, W.: Productivity loss in idea-generating groups: Tracking down the blocking effect. J. Pers. Soc. Psychol. 61, 392–403 (1991)
5. Gallupe, R.B., Bastianutti, L.M., Cooper, W.H.: Unblocking brainstorms. J. Appl. Psychol. 76, 137–142 (1991)
6. Wang, H.-C., Fussell, S.R., Setlock, L.R.: Cultural Difference and Adaptation of Communication Styles in Computer-Mediated Group Brainstorming. In: SIGCHI Conference on Human Factors in Computing Systems (CHI 2009), Boston, MA (2009)
7. Carroll, E.A., Latulipe, C., Fung, R., Terry, M.: Creativity factor evaluation: towards a standardized survey metric for creativity support. In: ACM Creativity and Cognition (C&C 2009), pp. 127–136. ACM Press, Berkeley (2009)
8. Torrance, P.E.: Torrance Tests of Creative Thinking. Norms-Technical Manual. Ginn and Company, Lexington (1974)
9. Chrysikou, E.G., Weisberg, R.W.: Following the wrong footsteps: Fixation effects of pictorial examples in a design problem-solving task. J. Exp. Psychol.-Learn. Mem. Cogn. 31, 1134–1148 (2005)
10. Nijstad, B.A., Stroebe, W.: How the group affects the mind: A cognitive model of idea generation in groups. Pers. Soc. Psychol. Rev. 10, 186–213 (2006)
11. Volkema, R.J.: Problem Formulation In Planning And Design. Manage. Sci. 29, 639–652 (1983)
12. Shah, J.J., Vargas Hernandez, N., Smith, S.M.: Metrics for measuring ideation effectiveness. Design Stud. 24, 111–134 (2003)
13. Perttula, M., Sipilä, P.: The idea exposure paradigm in design idea generation. J. Eng. Design 18, 93–102 (2007)
14. Kudrowitz, B.M., Wallace, D.R.: Assessing the Quality of Ideas from Prolific, Early-Stage Product Ideation. In: ASME DETC 2010, Montreal (2010)
15. Visser, W.: Design: one, but in different forms. Design Stud. 30, 187–223 (2009)
16. Vargas Hernandez, N., Shah, J.J., Smith, S.M.: Cognitive Models of Design Ideation. In: Design Engineering Technical Conferences (ASME DETC 2007), Las Vegas, NV (2007)
17. Liikkanen, L.A., Perttula, M.: Inspiring design idea generation: Insights from a memory-search perspective. J. Eng. Design 21, 545–560 (2010)
18. Shah, J.J., Kulkarni, S.V., Vargas Hernandez, N.: Evaluation of Idea Generation Methods for Conceptual Design: Effectiveness Metrics and Design of Experiments. J. Mech. Design 122, 377–384 (2000)
19. Cross, N.: Expertise in design: an overview. Design Stud. 25, 427–441 (2004)
20. Dow, S.P., Fortuna, J., Schwartz, D., Altringer, B., Schwartz, D.L., Klemmer, S.R.: Prototyping Dynamics: Sharing Multiple Designs Improves Exploration, Group Rapport, and Results. In: CHI 2011, Vancouver, BC (2011) (in press)
21. Mayer, R.E.: Fifty Years of Creativity Research. In: Sternberg, R.J. (ed.) Handbook of Creativity, pp. 449–460. Cambridge University Press, Cambridge (1999)
22. Srivathsavai, R., Genco, N., Hölttä-Otto, K.: Study OF Existing Metrics Used in Measurement of Ideation Effectiveness. In: ASME DETC 2010, Montreal (2010)
23. Yang, M.C.: Observations on concept generation and sketching in engineering design. Research in Engineering Design 20, 1–11 (2009)
24. Reinig, B.A., Briggs, R.O., Nunamaker, J.F.: On the measurement of ideation quality. J. Manage. Inf. Syst. 23, 143–161 (2007)

25. Shepherd, M.M., Briggs, R.O., Reinig, B.A., Yen, J., Nunamaker, J.F.: Invoking social comparison to improve electronic brainstorming: beyond anonymity. J. Manage. Inf. Syst. 12, 155–170 (1995)
26. Stroebe, W., Diehl, M., Abakoumkin, G.: The Illusion of Group Effectivity. Pers. Soc. Psychol. Bull. 18, 643–650 (1992)
27. Dugosh, K.L., Paulus, P.B., Roland, E.J., Yang, H.C.: Cognitive stimulation in brainstorming. J. Pers. Soc. Psychol. 79, 722–735 (2000)
28. Jansson, D.G., Smith, S.M.: Design fixation. Design Stud. 12, 3–11 (1991)
29. Purcell, A.T., Gero, J.S.: Design and other types of fixation. Design Stud. 17, 363–383 (1996)
30. Smith, S.M., Ward, T.B., Schumacher, J.S.: Constraining Effects Of Examples In A Creative Generation Task. Mem. Cogn. 21, 837–845 (1993)
31. Liikkanen, L.A., Björklund, T., Hämälainen, M.M., Koskinen, M.: Time Constraints in Design Idea Generation. In: International Conference on Engineering Design, ICED 2009, Palo Alto, CA (2009)
32. Oulasvirta, A., Kurvinen, E., Kankainen, T.: Understanding contexts by being there: case studies in bodystorming. Pers. Ubiquit. Comput. 7, 125–134 (2003)