Proceedings of the National Conference
On Undergraduate Research (NCUR) 2013
University of Wisconsin La Crosse, WI
April 11 – 13, 2013

Evaluating The Alternative Uses Test of Creativity

Caitlin Dippo
College of Design
University of Minnesota Twin Cities
100 Church St. SE
Minneapolis, MN 55455

Faculty Advisor Barry Kudrowitz

Abstract

The Alternative Uses Test is a means of evaluating divergent thinking abilities. The test requires subjects to list non-obvious uses for a common object. In our studies, participants were given three minutes to take this test using a paper clip as the object. There are several objectives of this research. Firstly, we are testing the hypothesis that the subjects who list more responses will have more creative responses. Similarly, we hypothesize that as one lists alternative uses, the items suggested first will be less creative/novel than the items towards the end. This data should provide insight into how people make associations while generating ideas. A secondary goal of this research is to develop an automated scoring rubric for this specific Alternative Uses Test to allow for a faster evaluation of creativity.

Keywords: Creativity, Alternative Uses, Paperclip

1. Introduction

Divergent thinking is a critical part of a creative design process. Design problems typically do not have one correct solution and often there are many solutions for a given problem. It is important to better understand how divergent thinking is evaluated and how people make associations to think of novel ideas. The alternative uses test is a standard test of divergent thinking. In the current alternative uses test manual from 1960, participants are asked to think of a fixed number of ideas (six) and those ideas are used in scoring. In this study, we removed the quantity constraint and participants were asked to list as many alternative uses for a paperclip as they can in three minutes to see when the novel ideas are generated. This test can be used to understand the relationship between quantity of ideas and novelty of ideas. We can also see how many ideas people produce before arriving at highly novel ideas. Novelty (or originality) is a crucial constituent of creativity and so in this study we will be focusing on the metric of novelty as measured by statistical originality. There is limited data on evaluating the alternative uses test and so this study is an opportunity to develop an updated and detailed scoring rubric for future use of this test as a measure of creativity.

2. Background

Creativity is a critical part of the design process. It takes creativity on the part of the designer to address the problems in new ways to develop novel solutions. It is the creative element that is the less common, less taught, less understood, yet more desired and influential aspect of design [1,2].

Nobel laureate Linus Pauling said, "…you aren't going to have good ideas unless you have lots of ideas [3]." Research has supported this notion with a positive correlation between total number of ideas and total number of good ideas [4,5]. Another study found that quantity of ideas was positively correlated with original ideas and negatively correlated with feasible ideas [6]. This work is building off a prior study that found that quantity of ideas in a brainstorming session was correlated with overall evaluation of creativity of ideas [7].

Osborn claims that the "early ideas are unlikely to be the best ideas generated during an ideation session" [8]. The first ideas we generate for a given prompt are going to be the common ideas that everyone first generates for that prompt [9]. The higher quality ideas are ones that have been built on earlier ideas [10,11].

The relationship between quantity and quality is not necessarily linear. One suggested relationship between idea quality and idea quantity is visualized with a curve called the ideation function [12]. The Bounded Ideation Theory (BIT) describes the ideation function as a positive s-curve; the number of good ideas gradually increases as the problem is better understood and, as participants get exhausted, the number of additional good ideas decreases [12].

For this study, to specify what is meant by quality ideas, good ideas, or creative ideas, we are referencing the "Standard Definition of Creativity." This definition of creativity is bipartite in which originality (or novelty) is the fundamental requirement and there is also a second factor, effectiveness, which often takes on different forms based on what is being evaluated [13]. Novelty is the only consistent requirement for creativity and so that is what is evaluated in this study.

The relationship between quantity and novelty is explored for a standardized test of creativity, the alternative uses test. If quality of ideas improves with quantity, how many ideas are needed to reach the novel ideas?

The majority of experimental creativity studies are based on creativity tests as they are relatively simple to administer and the data can be analyzed objectively. A few of the well known creativity tests include the Torrance Test of Creative Thinking (TTCT) by Paul Torrance [14], the Remote Associates Test (RAT) by Sarnoff Mednick [15] and the Guilford's Structure of the Intellect (SOI) divergent production tests by Joy Paul Guilford [16]. These tests measure certain cognitive abilities that have been shown to correlate with creative thought processes such as divergent thinking. A recent study found that divergent thinking is a reliable indicator of creative potential [17].

The alternative (or unusual) uses test, created by J.P. Guilford in 1967 as part of his Structure of Intellect (SOI), is a simple way of evaluating divergent thinking ability or, in his own words, "spontaneous flexibility" [16]. In this test, participants are asked to list non-obvious uses for a common object (such as a brick or a newspaper) in a fixed amount of time [16]. The responses are evaluated on 4 components: originality (statistically uncommon when compared to responses to the overall data set), fluency (quantity), flexibility (number of different categories), and elaboration (amount of detail). This scoring system is the basis of other creativity tests including the Torrance Test of Creative Thinking (TTCT) [14]. Guilford was the first to propose that it is possible to study and evaluate creativity of subjects using a psychometric approach with pencil and paper [18]. His tests began the usage of divergent thinking tests as the main instrument for measuring creativity [18]. In this study, we are using this test to:

1) Find relationships between quantity of ideas and novelty of ideas.
2) Gain insight into the way people make associations and evolve novel ideas from a common starting point.
3) Develop an updated and detailed scoring rubric for future use of this test as a measure of creativity.

3. Experiment

In this study, participants were asked to "list as many alternative uses for a paperclip as they can think of in three minutes." On a blank notecard, each participant wrote his or her gender, age, and responses. Half of the participants in this study are designers/engineers at large corporations and half are design/engineering university students. From a pool of over 2000 participants that completed the test, 293 were randomly selected and evaluated. There were 49 males and 121 females (123 participants did not give their gender) with ages ranging from 15 to 64 with a mean age of 28.

4. Results

All responses were digitized. In this study, the term response will refer to each alternative use that was listed on an index card. In total there were 2999 responses from the 293 participants. The median number of responses per participant was 10 and the average was 10.2 with a standard deviation of 4.6. A distribution of the quantity of responses is shown in Fig. 1.

The order in which each participant listed each response was preserved in the digitization. The only personal information recorded for each participant was their age, gender, and their industry/school affiliation.

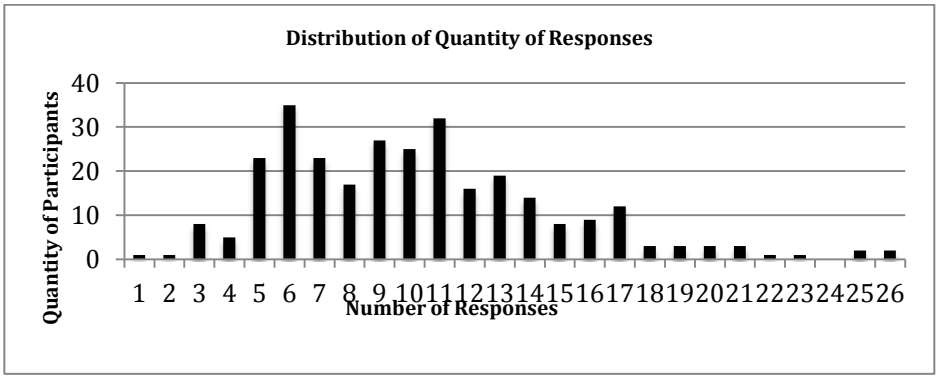

Figure 1. Distribution of Quantity of Responses

5. Categorizing Responses

We needed to create a system of classification in order to analyze the 2999 responses. Responses were digitized and also generalized into a single keyword that would simplify similar answers. Responses like "toothpick" and "remove food from teeth" were simplified into "tooth" because they both described the same use. Keywords group ideas that are the same but worded differently. In this study, we found 214 unique keywords that represent 214 uses for a paperclip.

Keywords were then further simplified into generalized functions that encompass many keywords. Keywords like "tooth" and "fingernail" were grouped together by the function of "removing material from small spaces". These functions map to the flexibility category for evaluating the test and so responses that are within one generalized function would not show flexibility in thinking. 80 unique generalized functions were found.

Generalized functions were even further simplified into categories of paperclip treatments of how the paper clip was being used. "Removing materials from small spaces" and "puncturing through something" would both be categorized under using a paper clip as a "pick". We have found 8 treatments of the paperclip that include clip, loop, pick, flexible wire, straight rod, flat token, material property, and abstract/artifact/symbol.

Once we had documented the responses (and keywords), we were then able to determine statistical originality of each response. We counted the number of times each response was used as well as the percentage of participants that gave each response. The less common a response (or keyword) is used by participants the more original the response. In this study we were solely looking at originality, as it is a good predictor of creativity [13], however, responses that did not make sense in the context or were illegible were omitted.

A response that was deemed original would have a *low percent of occurrence* in the participant pool for example "test a cake for doneness" had about 1% occurrence. Oppositely, an unoriginal response would have a *high percent of occurrence.* For example "hold paper together" was the most common response with a 77% occurrence. Figure 3 shows the most common keyword responses which are those given by over 10% of the participants in this study.

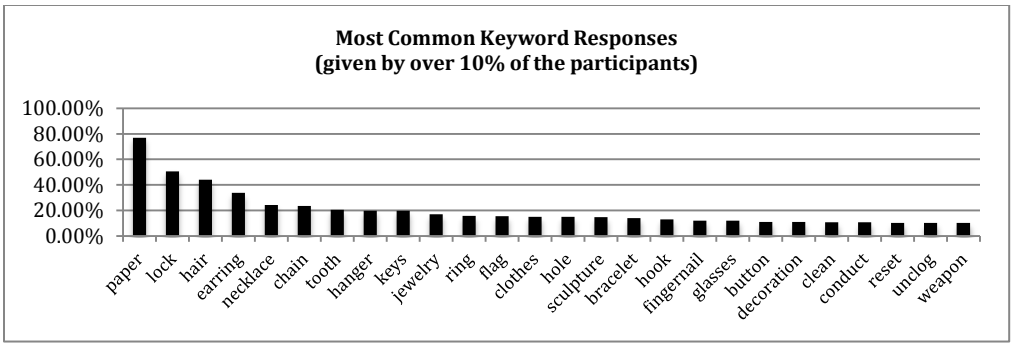

Figure 3. The most common keyword responses

Figure 3 would normally extend outward with a very long tail representing over 214 different alternative uses for a paperclip, but we cut the graph off at 10% occurrence. The long tail refers to the large number of uses/responses that are only mentioned by a few or one of the participants.

Similar to design problems, a divergent thinking test has many appropriate solutions.

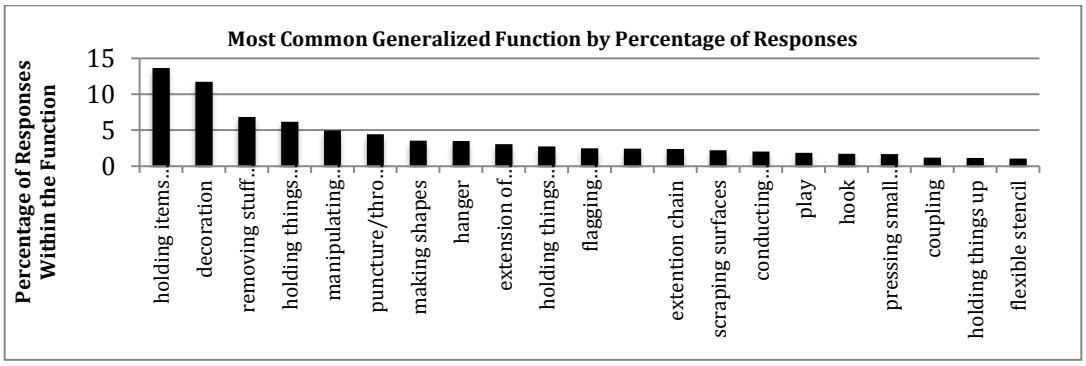

Figure 4. The most common generalized functions

When grouping the responses into generalized function, we can create a similar visualization for the most common generalized function by percentage of responses. Figure 4 would also continue outward with a long tail representing 80 unique generalized functions. The functions shown in the graph are those that encompass more than 1 percent of the responses.

Finally, we can visualize the percentage of the total responses using the high order of classification, clip treatment. As shown in Fig. 5, the most common treatments for the paperclip were as a pick, a clip, a flexible wire, and a loop. These four treatments represent 86% of all responses.

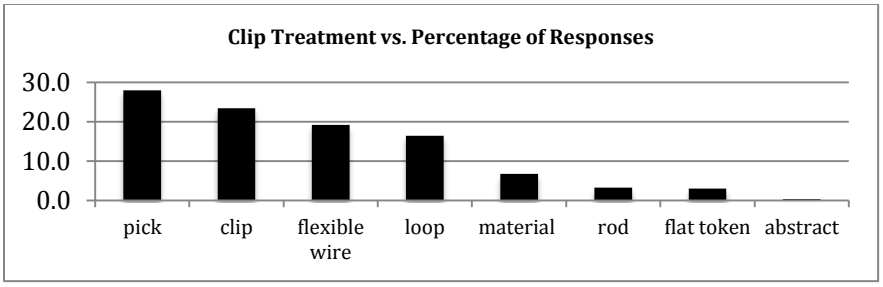

Figure 5. Comparing the occurrence of the 8 treatments

6. Quantity and Originality

We are interested in if originality of responses was related to the quantity of responses as prior studies have found a positive correlation [3,4,7]. Figure 6 and Figure 7 below show that participants that listed more responses had a lower average percent occurrence for their responses and thus more novel responses. An average percent occurrence of .1 means that on average 10% of the sample population also thought of the same ideas.

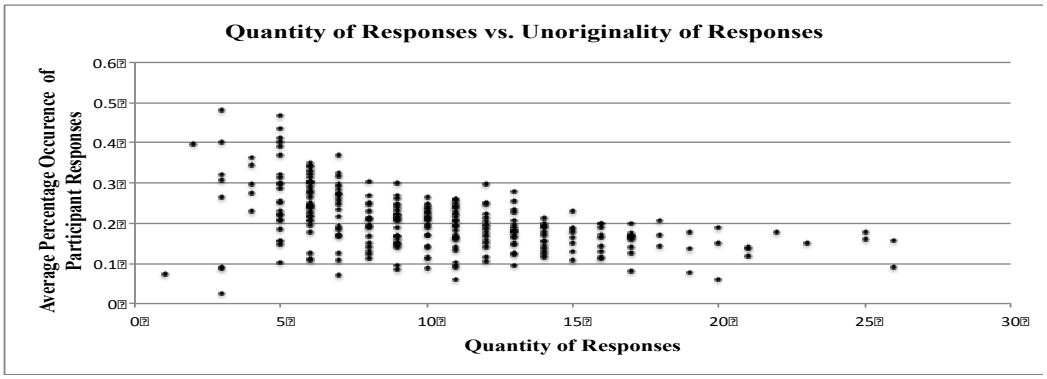

Figure 6. Quantity of Unoriginality vs. unoriginality of responses of each participant

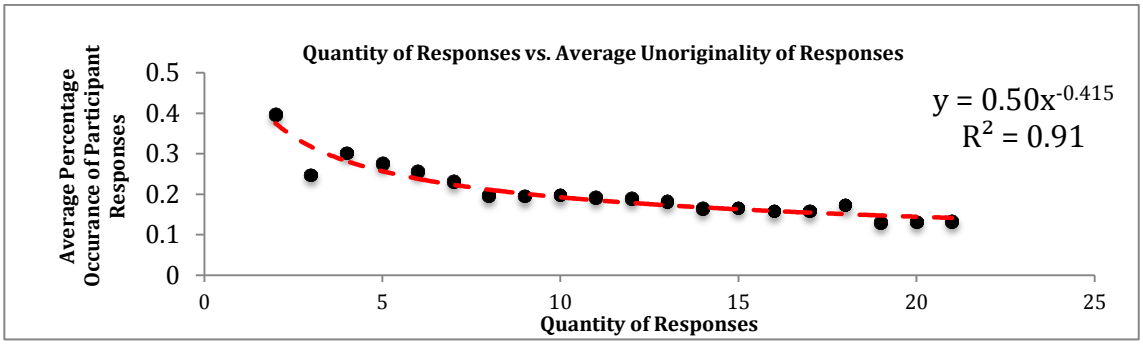

Figure 7: Quantity vs. average unoriginality of responses

To explore this further, we wanted to see if the later responses were more original than the former. By recording the order in which participants wrote down responses, we averaged the occurrence percentage for all responses written first, second, third, *etc*.

As shown in Fig. 8, the average occurrence percentage for the responses that were listed first was 47%. This means that if you took this test, it is likely that about half the people taking the test will have also written down whatever response you wrote first. Furthermore, for a given participant, later responses were significantly more novel than early responses and unoriginality of responses exponentially decreased with quantity at a rate of $x^{-1/2}$ (r^2=.94). Another way of saying this is that as the quantity of ideas goes up, the originality of those ideas goes up approaching a limit.

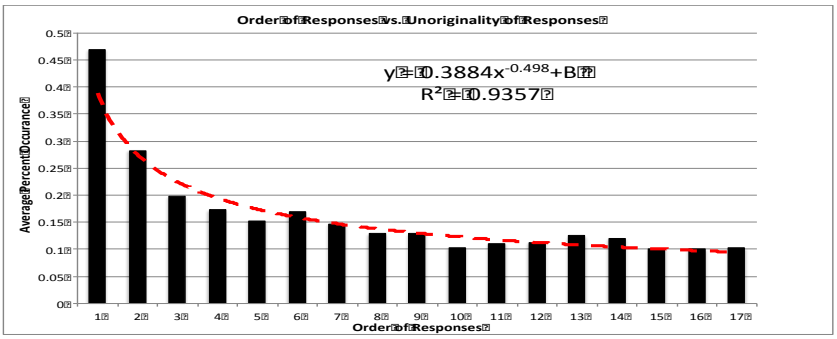

Figure 8: Order vs. unoriginality of responses

On average, a participant would list 9 responses before arriving at responses that were thought of by less than 10% of the participant pool. Participants that did not reach 9 responses in the study were likely to have few if any of these types of less common responses. With participants giving an average number of 10 responses in three minutes, we can see that the last responses a participant provides will have approximately a 10% occurrence percentage. Again, this means that if you write down 10 alternative uses, only 10% of the participant population will have also thought of your 10th response.

Another way of visualizing the data is by plotting the total number of novel ideas in the sample pool as a function of the number of responses. This is a means of comparing the data in this study to the Bounded Ideation Theory. As more ideas are generated, more novel ideas are produced until a point where the number of new novel ideas begins to level off. Figure 9 shows this phenomenon with varying the definition of novel ideas as those that are less than a 10%, 5% and 1% occurrence.

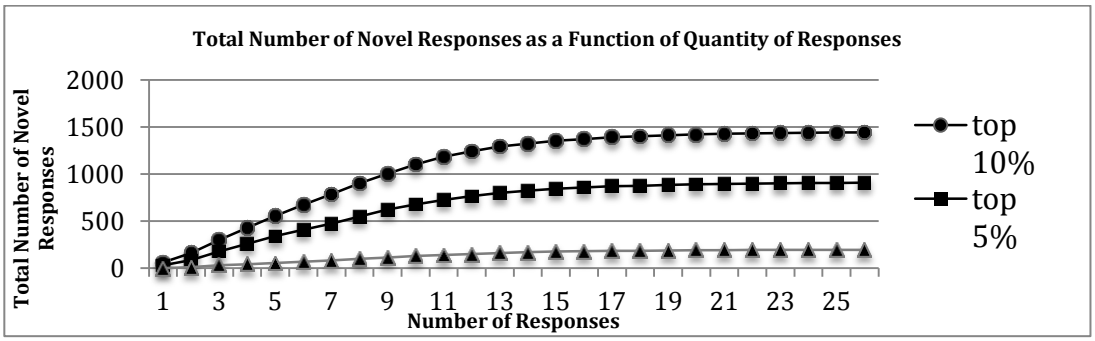

Figure 9. Total number of novel ideas as a function of quantity of responses

7. Discussion

The alternative uses test is only a measure of divergent thinking ability and so it may not fully represent general creativity. This data is not taking into account usefulness of responses (only appropriateness) and so we will use the term "novel ideas" as opposed to "creative ideas" to avoid confusion. This test, however, is a reliable indicator of creative potential [17]. It is also a standard test of creativity that is physical in nature as opposed to language based (like the remote associates test or parts of the torrance test of creative thinking).

Using percentage of occurrence as measure of novelty, it was found that participants that produced more responses had more novel responses and a higher average novelty score. As shown in Fig. 8, we found that the unoriginality of ideas decreases exponentially at a rate of x-1/2 (r2=.94). Participants listed approximately nine ideas before arriving at ideas thought of by less than 10% of the participant group. This means that if participants were given the current Alternative Uses Test, they would never reach the more creative ideas because they would be stopped at a maximum of six ideas.

If the Alternative Uses Test maps to real world problem solving, it suggests that the first handful of ideas we think of are likely to have been suggested already by others and thus not original. To get more novel solutions, one must

push past and build upon the ideas generated first to arrive at the less obvious ideas and associations. In this specific test, nine ideas marks the limit before the ideas begin to be highly novel, future studies should explore if this number applies to other problems and prompts.

We can also view this data in terms of the Bounded Ideation Theory through plotting the total number of novel responses as a function of quantity of responses (shown in Fig. 9). The S-curve is somewhat visible especially when using 5% and 10% occurrence as the definition of a novel idea. This graph shows that the number of good ideas increase as people have more time to understand the problem/prompt. As more ideas are produced, cognitive overload and physical exhaustion begin to take effect and number of new novel ideas begins to decline [12]. This results in the leveling off of the number of novel ideas.

There are few resources available for scoring the alternative uses test. There is one rubric available for online purchase but it does not include data for a paperclip [19]: http://www.mindgarden.com/products/altu.htm#ms

The manual's grading criteria are based on whether or not the listed use is possible. If it is, participants are given a point and if responses are impossible or too vague, participants receive no points [19]. In this study, the abstract responses tend to be the least common responses and therefore the most novel. It would be an oversight to dismiss the abstract alternative uses.

From this study, we developed a database of the most common responses, the most common general categories and higher order categories of the paperclip prompt. These databases can be used to measure originality and flexibility which are two of the four metrics used to evaluate the alternative uses test. The other two metrics fluency (quantity) and elaboration (detail) do not require a database for scoring.

8. Conclusions and future work

The alternative uses test is a standard test of divergent thinking. Using statistical rarity of responses as a measure of novelty, it was found that participants that produced more responses had more novel responses and a higher average novelty score. Furthermore, for a given participant, later responses were significantly more novel than early responses and unoriginality of responses exponentially decreased with quantity at a rate of $x^{-1/2}$ (r^2=.94). On average, a participant would list 9 responses before arriving at highly novel responses (thought of by <10% of the participant pool). The data supports the Bounded Ideation Theory in which the number of novel ideas increases at the beginning of idea generation and then level off as participant become fatigued forming an S-shaped curve. If this test maps to real world problem solving, it suggests that the first handful of ideas we think of are likely to have been suggested already by others and thus not original. To get more novel solutions, one must push past (and build upon) the ideas generated first to arrive at the less obvious ideas and associations.

This data also suggests an improvement to the standardized Alternative Uses Test. If participants are allowed to generate as many ideas as they can in a fixed amount of time as opposed to limiting them to 6 ideas, the participants are more likely to reach the more novel ideas. A second suggestion for improvement to the standardized Alternative Uses Test is to not discard the abstract ideas as those were found to be the most statistically original and thus most novel.

In the future, this test could be digitized which would allow for real time scoring. The software could determine the appropriate keyword for each response. An algorithm could also be created to predict next responses based on ideas already given. This could be used to visualize the way people make connections and how novel ideas evolve from a common starting place.

Other directions to explore with this data are the relationship between flexibility and novelty. Are participants who jump between categories producing more novel responses or just the obvious responses from each category? Is there a correlation between elaboration and novelty? The current database made for this study does not consider the elaboration component as we have condensed similar ideas with different levels of detail into a simple keyword. One can refer back to the original responses and determine if participants who are verbose with their responses are also producing original responses. In this study, abstract responses (trade it, sell it, *etc*.) were the least common of the paperclip treatments and this would mean they are statistically the most original responses. Are the participants who produced abstract responses also scoring highly on other metrics such as originality and fluency.

We plan to continue entering data from our sample pool of participants to ensure that the general trends and correlations remain true.

9. Acknowledgements

The authors wish to express their appreciation to the University of Minnesota for funding this research and the Office for Information Technology Tech Stop for donating their time to help with the organization of the response database.

10. References

[1] Nussbaum, B., Berner, R., & Brady, D., 2005. "Get Creative! How to Build Innovative Companies". *Bloomberg Buisnessweek*, 1 August.
[2] Pink, D., 2005. *A Whole New Mind: Why Right-Brainers Will Rule the Future*. Riverhead Books, NY.
[3] Pauling, L., 1977. Interview. In: *Linus Pauling, Crusading Scientist.* TV, WGBH-Boston.
[4] Diehl, M., & Stroebe, W., 1987. "Productivity loss in brainstorming groups: Toward the solution of a riddle". *Journal of Personality and Social Psychology, **53***(3), pp. 497-509.
[5] Paulus, P, Kohn, N, & Arditti, L., 2011. "Effects of the Quantity and Quality Instructions on Brainstorming". *The Journal of Creative Behavior*, **45**(1), pp. 38-46.
[6] Rietzschel, E. F., Nijstad, B. A., & Stroebe, W., 2006. "Productivity is Not Enough: A comparison of Interactive and Nominal Groups in Idea Generation and Selection". *Journal of Experimental Social Pschology,* **42**, pp. 244-251.
[7] Kudrowitz, B. & Wallace, D., 2012. "Assessing the Quality of Ideas from Prolific, Early Stage Product Ideation". *Journal of Engineering Design: Special Issue on Design Creativity*. ifirst, 1-20.
[8] Osborn, A., 1963. *Applied Imagination: Principles and Procedures of Creative Problem Solving.* Charles Scribners Sons, New York, NY.
[9] Johnson, S., 2010. *Where Good Ideas Come From.* Riverhead Hardcover, New York, NY.
[10] van der Lugt, R., 2001. "Relating the quality of the idea generation process to the quality of the resulting design ideas" in *A Folkeson, K Gralen, M Norell and U Sellgren (eds) Proceedings of 14th International Conference on Engineerin Design Society*, Stockholm.
[11] Goldschmidt, G. & Tatsa, D., 2005. "How good are good ideas? Correlates of design creativity". *Design Studies*, **26** (6), pp. 593-611.
[12] Reinig, B.A., R.O. Briggs, and J.F. Nunamaker., 2007. "On the measurement of ideation quality". *Journal of Management Information Systems*, **23** (4), pp. 143-161.
[13] Runco, M. & Jaegar, G., 2012. "The Standard Definition of Creativity". *Creativity Research Journal,* **24** (1), pp. 92-96.
[14] Torrance, E.P., 1972. "Predictive Validity of Torrance Tests of Creative Thinking". *Journal of Creative Behavior*, **6**(4), pp. 236-252.
[15] Mednick, S. A., 1962. "The Associative Basis of the Creative Process". *Psychological Review,* **69** (3), pp. 220-232.
[16] Guilford, J.P. 1956. "The Structure of Intellect". *Psychological Bulletin,* **53**(4), pp. 267-293.
[17] Runco, M. & Acar, S., 2012. "Divergent Thinking as an Indicator of Creative Potential". *Creativity Research Journal*, **24** (1), pp. 66-75.
[18] Sternberg, R.J. and Lubart, T.I., 1999. "The Concept of Creativity: Prospects and Paradigms*"* in *Handbook of Creativitity*, R.J. Sternberg, Editor., Cambridge University Press, New York.
[19] Guilford, J.P., Christensen, P.R., Merrifield, P.R., and Wilson, R.C., 1960. *Alternative Uses Manual.* Sheridan Supply Co.