Assessing the quality of ideas from prolific, early-stage product ideation

Kudrowitz, Barry Matthew, and David R. Wallace, “Assessing the quality of ideas from prolific, early-stage product ideation,” ASME 2010 International Design Engineering Technical Conferences and Computers and Information in Engineering Conference (EIDTC 2010), August 15–18, 2010, Montréal, Québec. (New York: ASME, 2010): p. 381-91 doi 10.1115/DETC2010-28991 ©2010 Authors

As Published
http://dx.doi.org/10.1115/DETC2010-28991

Publisher
ASME International

Version
Final published version

Citable link
https://hdl.handle.net/1721.1/124409

Terms of Use
Article is made available in accordance with the publisher’s policy and may be subject to US copyright law. Please refer to the publisher’s site for terms of use.
ASSESSING THE QUALITY OF IDEAS FROM PROLIFIC, EARLY-STAGE PRODUCT IDEATION

Barry Matthew Kudrowitz  
Massachusetts Institute of Technology  
Department of Mechanical Engineering  
Cambridge, MA 02139

David R. Wallace  
Massachusetts Institute of Technology  
Department of Mechanical Engineering  
Cambridge, MA 02139

ABSTRACT
The goal of many idea generation techniques, such as brainstorming, is to generate a large quantity of ideas with the hopes of having a few outstanding novel ideas that are worth pursuing. The output of such sessions is a large number of rough concept sketches, which require a rapid means of screening to select a manageable set of promising ideas. In this study we develop and test metrics for evaluating large quantities of early-stage product idea sketches. In total, 1767 ideas for three different product themes were used as a test bed. With our findings, we suggest three independent qualities that fully describe an innovative product idea: creative (as a subjective judgment), useful (as defined as having practical applications), and feasible (as determined by experts). Reviewers’ subjective ratings of idea creativity had a strong correlation with ratings of idea novelty (r²=.80), but negligible correlation with idea usefulness (r²=.16). The clarity of sketch positively influenced ratings of idea creativity. Another interesting finding is that the quantity of ideas generated by the individual subjects had a strong correlation with that subject’s overall creativity scores (r²=.82) and novelty scores (r²=.85), but had weak correlations with that subject’s usefulness scores (r²=.38).

INTRODUCTION
Product design idea generation (e.g. brainstorming) results in a large quantity of concept sketches which require a rapid screening process to select the most innovative. There are many studies that evaluate the creativity of ideas and they rarely use the same instrument of metrics for rating [1]. The majority of these studies are specific to ideas for situational problems and not products. The few metrics that are intended for products are specific to finished artifacts and not ideas for products [2-15]. In addition, these product rating metrics are designed for reviewing a few products in detail and not hundreds of ideas. This research tests a simple metric for quickly rating the innovativeness of a large quantity of product ideas. This study focuses on evaluation of sketches of ideas, as it can be applied to the visual thinking methods widely used in product design firms such as IDEO.

IDEA GENERATION
There are many methods that designers use to come up with new ideas for products and creative solutions to problems. These idea generation (or ideation) methods can be grouped into categories in several ways.

One means of classifying methods is by the type of problems they are addressing. Problems can be ill structured when they have poorly defined goals [16] such as “we need new types of water toys.” An example of an ill structured task would be designing a house where there are no initial definite criterion to test a solution [17]. These types of problems often have a large number of solutions and there is no singular best answer. Problems that are structured, such as “find the least expensive means of securely hanging this picture frame,” have well defined goals and often there is an optimal solution. As ill structured tasks begin to acquire structure and direction they can be decomposed into many well structured problems [17].

Just as there are ill structured (or free form) and structured problems, there are also freeform and structured idea generation methods. Freeform idea generation methods include: brainstorming, free association and brain writing; and structured idea generation methods include TRIZ, morphological analysis, and SCAMPER [18, 19]. The freeform methods tend to favor blue-sky or early-stage design (typically performed as a group), while the structured methods are more suitable for more advanced stages of design after a general concept has been selected (typically performed by individuals). This study relates best to free form ideation methods and, in particular, brainstorming.
Blue-sky idea generation as well as product design in general tends to favor innovation over invention. Invention is creating a new and useful technology. Innovation is the combination of knowledge or technologies in original and non-obvious valued new products, processes or services [20]. The contemporary design engineer is expected to be able to participate in the creative early stages of design in addition to the more technical elements [21].

Brainstorming
In a brainstorming session, a group of people sits together to generate a large number of ideas in a short amount of time [22]. The participants quickly sketch any idea that comes to mind and present it to the group. Each sketch should be drawn large and clearly on a separate sheet of paper and given an appropriate and simple title. As ideas are generated, a group facilitator calls on members to share their ideas. The group facilitator ensures everyone hears each idea and then records the concept title with a brief description.

The four main rules of a good brainstorming session are to defer judgment and not critique ideas, build off of each other’s ideas, encourage wild ideas and come up with as many ideas as possible [22]. All participants must be able to speak freely without fear of being subjected to harassment. No ideas are to be considered “silly” or discarded as infeasible, as these irrelevant or fanciful ideas may lead to plausible concepts.

Nominal Brainstorming
In a nominal brainstorm session, the participants do not talk to each other. They either share their ideas silently or do not share the ideas until after the session ends. Nominal brainstormings are used in studies, such as this one, when individual ideation ability is of interest. It has also been argued that nominal brainstorming is more effective than traditional brainstorming as it eliminates three theoretical factors of productivity loss: production blocking, evaluation apprehension and free riding [23]. Production blocking is the term given to the phenomenon that occurs when participants in a brainstorm session suppress or forget their ideas because they seem less relevant and original or other speaking members distract them. Evaluation apprehension is the term for the fear of negative evaluation from other group members. Free riding is the term for participants not contributing to the group as they feel that their contributions are dispensable. As the number of group members increases, the number of free riders, or non-participants, also increases [24].

PRIOR STUDIES OF IDEA RATING
Between 1990 and 2005, there have been approximately 90 studies that have dealt with analyzing ideas [1] and they rarely use the same instrument of metrics for rating. The majority of these studies involve prompts related to situational problems (e.g. increasing tourism in Tucson, improving the psychology department at a college, or dealing with a drug dealing roommate). Ideas for situational problems are quite different than ideas for products as the latter typically involve sketching and artifacts. Situational prompts are also highly structured where there can be “more correct” responses and usefulness has a large impact on the perceived goodness.

The few studies that are specific to rating products typically involve existing finished artifacts and not ideas [2-15]. These studies also review a few products and not hundreds. The work of Susan Besemer [4] is perhaps the only work that deals with ideas for products, but is more suited for reviewing and comparing only a few ideas in detail. A hybrid of these instruments is needed to evaluate product ideas from a high volume product ideation session.

Metrics for Evaluating Ideas
The “criteria problem” is the term given to the issue of obtaining a valid assessment of the level of creativeness of a person [7]. In the studies that have dealt with product ratings, there are various sets of criteria that have been suggested as to what determines a creative product.

There are two main categories of definitions for creative products: novelty-based and multi-attribute based [1].

Multi-attribute based definitions of creativity claim that products must be novel as well as possess other quality attributes. The most commonly used quality attribute to describe a creative product is usefulness [1]. Mednick defines creative thinking as “…the forming of associative elements into new combinations which either meet specified requirements or are in some way useful [25]” Other quality attributes include relevance, appropriateness, clarity, workability, feasibility etc. Almost every study in this field chooses different attributes to determine creativity. According to Shah and Vargas, good engineering design should be both practical and novel [13]; however, engineering design is not synonymous with product design.

Novelty-based definitions claim that products are creative based solely on originality and do not depend on any other quality factors such as appropriateness, usefulness or applicability. Runco and Charles found that “it is not necessary for an original idea to be appropriate (i.e solve a problem) to be viewed as creative [26].” In a more recent study, Christiaans found that “usefulness seemed not to be important in discriminating between designs with high and low creativity ratings [9]”. Ideas that are novel, but not feasible or useful are sometimes termed aesthetic innovation [27], chindogu [28], or novelty items.

Amabile takes a stance that is divergent from these attribute-based assessments and claims that creativity cannot be determined objectively using metrics. She suggests a subjective Consensual Definition of Creativity: “A product or response is creative to the extent that appropriate observers independently agree it is creative [3].” In other words, instead
of determining if a product is creative by asking reviewers if the product is novel, useful (etc.), simply ask reviewers if the product is creative.

Table 2 presents relevant studies and the dimensions used to determine the creativity of a product (idea or artifact).

Table 2. SELECTION OF PRIOR STUDIES INVOLVING A METRIC FOR THE EVALUATION OF IDEAS/PRODUCTS

| Study                                      | Summary                                                                 | Products in Review | Dimensions of Creativity                                                                 |
|--------------------------------------------|------------------------------------------------------------------------|--------------------|-----------------------------------------------------------------------------------------|
| Dean et al (2006)                          | Review of 90 Constructs for Idea Evaluation                           | Ideas (e.g. increase tourism in Tucson) | Novelty (originality, paradigm relatedness) Workability (acceptability, implementability) Relevance (applicability, effectiveness) Specificity (implicational explicitness, completeness and clarity) |
| Besemer and O'Quin (1986, 1999)*           | Objective Metric for Creative Product Evaluation                      | Artifacts (chairs) | Novelty (surprising, original) Resolution (logical, useful, valuable, understandable) Elaboration and Synthesis (organic, well-crafted, elegant) Quantity |
| Shah and Vargas-Hernandez (2000, 2002)     | Evaluation Of Mechanical Engineering Designs                          | Artifacts and Ideas (mechanical devices) | Novelty (unusual) Quality (meets specifications) Variety (explored solution space) |
| Horn and Salvendy (2006, 2009)*            | Consumer Based Assessment of Product Creativity                       | Artifacts (chairs and lamps) | Novelty (frequency, rarity) Importance (relevance, significance) Affect (appeal, desire, attraction, delight, stimulation, etc) |
| Amabile (1982)**                           | Subjective Assessment Method of Creativity                            | Artifacts (artwork and poetry) | Creativity (as determined by appropriate judges) Creativity Cluster (novel material use, novel idea, effort, detail, etc.) Technical Cluster (technical goodness, organization, neatness, etc.) Aesthetic Judgment (liking, aesthetic appeal, would you display it?) |
| Christiaans (2002)**                       | Creativity as one metric of design review                             | Artifacts and telephone booths | Creativity (as determined by appropriate judges) Technical Quality Attractiveness Interest Goodness of Example |

*In these studies, each of the dimensions are made up of many sub-dimensions in the form of bipolar adjective scales
**In these studies, the assessment dimensions in addition to “creativity” do not determine creativity, but were used to find correlations with creativity

Quantity of Ideas

The quantity (sometimes termed fluency or productivity) produced in an ideation session is a common dimension that is reviewed in addition to the dimensions that are used to measure creativity. For the most part, quantity is objective and easy to measure. In most studies, a count of non-redundant ideas constitutes the rating for this dimension.

Linus Pauling said, “The best way to have a good idea is to have a lot of ideas.” Research has shown a correlation between total number of ideas and total number of good ideas [23]. Another study found that quantity of ideas was positively correlated with original ideas and negatively correlated with feasible ideas [29]. Dean et al found that 18 of the 90 idea review studies between 1990 and 2005 used quantity as the sole means of evaluating the ideation session [1].

Some studies evaluate only one idea per person or group, which makes it easier for a reviewer to evaluate those few ideas on several dimensions. When there are many ideas in need of review, the number of dimensions for which each idea is evaluated must be limited to prevent reviewer fatigue. Amabile found that levels of interjudge agreement depend on the magnitude of effort required by the judges [3].

Judging Processes

The only requirement of judges is some basic level of familiarity with the subject matter [3]. Expert judges are only required when the products in review are of a highly specialized nature. “The more ‘cutting edge’ a product is in a specific domain, the more likely an expert judge will be required [30].” In the case of common household items (such as a toaster, umbrella or toothbrush), almost anyone could be considered an appropriate judge of creativity. A few studies have found, based on interrater agreement, that there is little difference between experts and non-experts in rating design creativity [3, 9]. Although professionals or experts are capable of producing creative products, they may be unreliable in judging them as they rely on higher, esoteric, or idiosyncratic standards [30].

In some experiments [29, 31] a panel rates dimensions (such as novelty and practicality) on a scale (e.g. from 1-5). If scores are to be combined for an individual’s set of ideas, they will either be added or averaged. If added, a large number of bad ideas are helpful and if averaged, a large number of bad ideas will lower the overall score. Some argue that it is best to simply count the number of creative (practical, creative, etc) ideas as opposed to average a rating to avoid the influence of a large number of bad ideas [32]. An argument against the count method is that it does not take into account the degree of the creativity of the ideas.

In some cases, the judges agree upon a standard for what a 1, 3 and 5 score would be or how to define each of the dimensions. When dealing with the dimension of creativity, Amabile suggests that the judges use their own definitions of creativity as opposed to attempting to define criteria [3].
EXPERIMENT

Selecting Ideation Themes
We chose to have the participants ideate around a product theme (ill structured) instead of developing solutions to a problem (structured). There are several reasons why we chose an ill structured or blue-sky idea generation over a structured theme. In structured problems, the solutions may not be related to products at all and it would then be hard to compare innovative designs. Designers and engineers may develop better solutions to structured problems as a result of their training in the scientific method and less because of creative abilities. The structured problem prompt also suggests there is in fact a solution or best solution and we are trying to avoid that type of thinking [16]. Finally, in our recent experiences, companies often come to us asking for blue-sky innovation around a product theme such as peristaltic pumps, water guns, or cork. These are realistic scenarios. In the future, it may also be beneficial to look into a structured problem ideation as this may involve a different thought process and a different form of creative thinking.

We chose umbrellas, toasters and toothbrushes as the three product themes for subjects to ideate. These themes are common enough for the general public to have a good understanding of the current state of the art. The themes are also common enough for the “average” person to be considered an appropriate judge of creativity [3]. A similar study evaluating the creativity of products chose two of these same themes (toasters and toothbrushes) because of their “commonality, moderate cost, and level of interest for general young professional consumers [12].” This similar study evaluated products that were currently on the market and not conceptual product ideas.

Participants
We administered this test to 84 participants (52 male, 32 female). The test was administered to our four interest groups: 24 product designers, 21 improvisational comedians, 26 MIT students, and 13 others. The combined age range of the participants was between 18 and 63 years. The mean age was 28 and the median age was 23. The MIT students that volunteered to participate were all involved in one of three different product design-related courses (2.00b Toy Product Design, 2.009 Product Engineering Processes, or 2.97 DesignApalooza). The majority of the students were underclassmen and more specifically freshmen.

Participants were given the test in familiar locations, typically in their work surroundings in a group at a large conference table. Tests were performed individually but sitting together as a group to save time on the part of the administrator and to make the participants more comfortable.

Procedure
We decided to use a nominal brainstorming opposed to a traditional interactive brainstorm so we can better evaluate the individual participants.

Participants were asked to develop as many innovative concepts they can around a given product theme. Each participant was given a stack of blank legal paper and a black fine tipped Sharpie® permanent marker. They were asked to sketch each idea with a title on a new piece of paper in the portrait orientation. We asked that all sketches be made in the portrait orientation to simplify the scanning, processing and reviewing. Participants were asked to draw large, legibly and use as few words as possible to explain the concept. They were told that drawing ability does not count and that they should sketch all ideas that they feel are innovative in some way. Participants were told explicitly to not talk to each other or share ideas.

Each group of participants ideated around all three themes in a different order. Each brainstorm session lasted 12 minutes. In a pretest we found that 12 minutes is an acceptable time to limit fatigue and still produce a good quantity of ideas. Time was called at 12 minutes and participants were allowed to finish any sketch they had started.

An identification number specific to each participant was placed on the backside of each concept drawing and each drawing was scanned.

Designing a Metric
In a smaller test experiment prior to the main study, there was confusion on the meanings of usefulness, practicality, feasibility and novelty and whether these criteria actually determined innovativeness or creativity.

In choosing appropriate evaluation metrics, we referenced the prior studies of Table 2. The classification presented by Dean et al. [1], based on MacCrimmon and Wager [33], is thorough, but is also tailored to evaluate ideas for themes that are outside the realm of products. Dean suggests reviewing ideas by the following classification: Novelty (composed of Originality and Paradigm Relatedness), Workability (composed of Acceptability and Implementability), Relevance (composed of Applicability and Effectiveness), and Specificity (composed of Implicational Explicitness, Completeness, and Clarity).

Horn and Salvendy [12], Besemer and Treffinger [7] are a few of the researchers that deal specifically with metrics for reviewing products. Some concerns with their metrics are that they deal with products that already exist. They are also better suited to compare a few designs as opposed to hundreds of concepts and so their metrics can be long and detailed.
In this study, we evaluated ideas based on quantity as well as five qualitative metrics: creative, novel, useful, product-worthy, and clear. Descriptions of these metrics are as follows:

Creative – Using the judge’s own subjective definition of creativity [3]. Following in the subjective assessment style of Amabile [3] and Christiaans [9], we decided to ask the reviewers to rate the ideas on several metrics, one of which is creativity as defined by the reviewer. We can then assess which metrics correlate with the subjectively rated creative products.

We also decided to have reviewers subjectively rate on “creativity” of the products rather than “innovativeness” of the products. Even though the definition of “creativity” is subjective, we believe that majority of people have a better internalized meaning of “creativity” as compared to “innovative.” It is possible to then take a more traditional objective definition of “innovative” such as the combination of knowledge in original and non-obvious valued new products, processes or services [20] which would equate to “novel” plus “useful” plus “product-worthy.”

Novel - The concept is original [8] and uncommon [12]. A gun that shoots out celery may not be considered useful or practical, but it should be considered novel. According to the multi-attribute definitions of creativity, novelty does not always imply creative. Take the celery gun example and imagine if the prompt was to ideate new toothbrush concepts. This concept would be considered original and uncommon, but would not make any sense in the realm of toothbrushes. As Mednick explains “7,363,474 is quite an original answer to the problem ‘how much is 12+12’” but it would not be considered creative [25].

There is said to exist three types of novelty: novel to the individual, novel to the society/group, novel to history [13]. An idea for a product that is only novel to the individual would most likely not be considered innovative. In the realm of product design, societal and historical novelty is more important than individual novelty.

Useful – The concept has practical applications in the product theme. This is perhaps the most controversial category for rating concepts. MacCrimmon and Wagner [33] and Dean et al. [1] would call this category “Relevance” meaning applicable the problem or topic at hand with an effective solution. This wording is more appropriate for structured prompts as they have solutions. Horn and Salvendy [12] call this category “Importance” defined as relevant and significant. Besemer and Treffinger [7] and Besemer and O’Quin [6] would call this category “Resolution” defined as the correctness of the solution; it is logical, useful and valuable. In this research study we decided to separate “value” from “use” as products can have value without being of practical use.

Product-worthy – The concept is both feasible and marketable. Feasibility can be described as the ability of the idea to be physically made into a product. It takes into account technology, cost, safety, and manufacturability. Marketable takes into account the social and cultural variables and if people would want to purchase the product (is it too much work to use? is it too large to store? would it be too expensive for the intended users?). This category is termed “Workability” by MacCrimmon and Wagner [33] and Dean et al. [1] and they define it as acceptability and implementability. This category is not addressed by Horn and Salvendy [12] and Besemer and O’Quin [6] as their studies review products that are already on the market.

Product-worthy may be hard to determine for the layman, as there are many ways of manufacturing a product and many different technologies that are unknown to the general public.

Clear – The concept is well communicated. This metric maps to what Dean et al. [1] and MacCrimmon and Wagner [33] term “Specificity” meaning clear, complete and explicit. As we are dealing with brainstorming, it is not expected that ideas should be elaborate and complete, however we are expecting ideas to be clear and detailed enough that the judges are able to understand the concept. This category also relates to the Besemer an O’Quin [6] category of “Synthesis” which includes organic, well-crafted, elegant and, in an earlier study, understandable [8]. Unlike “creative”, “novel”, “useful”, and “product-worthy,” “clarity” is independent of the concept and based solely on the idea presentation.

We were also interested in adding the metrics of “humorous, “playful,” “desirable,” and “I would buy this,” but given the quantity of product ideas to review, these additional metrics would add to the review fatigue.

Rating Ideas
To evaluate the product concepts, we chose to use an online website approach as opposed to a physical review form. An online review has several benefits. The reviewers can be located in many different locations and thus allows for a better general population sample. The reviewers can do the rating at their own convenience and in a comfortable setting. The data can be reviewed by hundreds of people as opposed to a select panel of individuals. The review data is also easy to collect and tabulate. As the product ideas are common it is acceptable to have a random laymen review panel [3].

We used Amazon Mechanical Turk as the means of collecting reviews. Amazon Mechanical Turk is a website (http://www.mturk.com) that allows any user to post tasks for any other user to complete. The tasks are termed “HITs” or “Human Intelligence Tasks,” the person posting the task is termed a “requester”, and the people completing the tasks are termed “workers.” Requesters post HITs and a monetary
payment for completing each HIT. In this study we only allowed workers to participate if they were located in the United States (based on IP address).

Each worker was paid $0.15 to rate 20 randomly selected product ideas. Each product idea was presented as a scan of the original sketch alongside radio buttons for rating each metric. An example of the online review page is shown in Figure 1.

Twelve different workers out of a pool of 397 rated each of 545 toaster ideas, 627 umbrellas ideas and 595 toothbrush ideas. The workers had an average age of 34.4. Approximately 38% were male.

In evaluating individual ideas, an average of review scores for each metric is acceptable. This would give each idea a set of average scores between 0 and 2. However, when evaluating a set of ideas (i.e. the creativity of a participant) we believe that a total count of 2s (yeses) is the most equitable measure to ensure that good ideas are not devalued by a large number of bad ideas [32]. A count of good ideas (e.g. number of ideas with a score over 1.0) is also an acceptable means of scoring the individual participants, however, the count of 2s gives a greater depth and finer resolution.

RESULTS

Interrelationships between Metrics

Comparing the average scores for each idea, Table 1 shows the interrelations between each metric rating.

|           | Clear | Creative | Novel | Product-worthy |
|-----------|-------|----------|-------|---------------|
| Creative  | .32   | .42      | .24   | .43           |
| Novel     | .24   | .14      | .80   | .07           |
| Product-worthy | .43 | .16      | .08   | .86           |

Useful and product-worthy ($r^2=.86$) as well as creative and novel ($r^2=.80$) are practically indistinguishable to reviewers. Clarity appears to be moderately correlated to product-worthy and useful ratings, and minimally correlated to creative and novel ratings. One can notice greater effects of clarity on creativity when comparing a set of similar ideas that are sketched with different levels of clarity.

Figures 2, 3 and 4 show the most relevant inter-metric relationships.
Quantity of Ideas

Quantity of ideas was highly correlated with overall individual creativity scores ($r^2=.82$) as well as quantity of creative ideas (i.e., ideas with average creative scores over 1.0) ($r^2=.64$). It may be argued that the high creative scores is an artifact of having a lot of ideas, however, the useful score is only minimally correlated with quantity of ideas ($r^2=.38$). The quantity of ideas compared to individual creative and useful scores are shown in Figures 5 and 6 respectively.

If we plot the quantity of ideas to the total average creativity score per subject, there is a negligible negative/no correlation ($r^2=.01$). This, however, is not an accurate measure as subjects are penalized for having bad ideas mixed with good ideas. Ten good ideas out of 100 is better than one good idea out of two, even though the former has a much smaller percentage of good ideas. It is the number of good ideas that matter, not the number of bad ideas.

DISCUSSION

Quantity of Ideas

As quantity of ideas is highly correlated with creativity of ideas ($r^2=.82$), it is reasonable to measure an individual’s creativity simply on fluency. This, however, does not accurately represent that individual’s ability to produce useful or feasible ideas.

It could be argued that individuals that produce a lot of ideas are better at divergent or associative thinking, which is often related to creative thought process [34, 35]. As more associations are made, the probability of reaching a creative idea increases [25]. One could also argue that individuals that are uninhibited will edit thoughts less, produce many ideas and their output should be less restrained and thus more creative. Oppositely, individuals that are logical thinkers may come up with more useful concepts, but will also be restrictive in their thought process producing fewer concepts in total.

Rating Product Worth

We defined product-worthy as both feasible and marketable. These elements can be mutually exclusive which made this category difficult for reviewers. Laymen reviewers had difficulty rating marketability. Several ideas were rated as not marketable, when the idea was for a product on the market.
example of this is the Miracle Toaster shown in Figure 7a, where 12/12 reviewers rated it as not product-worthy when in fact it is currently a product on the market sold by Fred Inc. Designed by Jason Amendolara, shown in Figure 7b.

Figure 7. TOASTER THAT PRODUCES HOLY IMAGES
(http://www.worldwidefred.com/holytoast.htm, 2006)

Laymen are also not appropriate raters of feasibility. A good percentage of reviewers rated a battery-powered toaster, a hand-cranked toaster and a cardboard box solar toaster as product-worthy.

As marketability is dependant on a variety of factors external from the concept itself, we suggest simplifying the category of product-worthy to feasible with the caveat that it is determined by expert judges (engineers or designers).

Concept Clarity and Creativity

The general correlation between clarity and creativity scores was low-moderate ($r^2=.32$). However, with the large number of ideas generated, there were many ideas that were reoccurring, presented by different participants. Using these sets of similar ideas, we can better see the effect of clarity of sketch on perceived creativity.

Out of the 545 toaster ideas, four were ideas for toasters that burn a holy image into the toast. Figure 8 shows the ratings of clarity and creativity for each of these toasters as measured by a count of 2s. As the clarity score increase, the creativity score increases even though these ideas are essentially the same. Another example of this phenomenon can be found in Figure 9.

These findings suggest the importance of basic sketching ability for engineers. Creative ideas can be overlooked when they are poorly sketched.

Figure 8. CREATIVITY AND CLARITY OF HOLY IMAGE PRINTING TOASTER IDEA SKETCHES

Figure 9. CREATIVITY AND CLARITY OF OPTICAL BURN-DETECTING TOASTER IDEA SKETCHES
A Map of Product Innovation

Innovation is the combination of knowledge or technologies in original and non-obvious valued new products, processes or services [20]. In essence, what makes a product idea innovative is a combination of novelty (or creativity), usefulness (or practical value), and feasibility. With strong correlations between creative and novelty scores ($r^2=.80$), we believe it is appropriate to rate either creativity or novelty as both would be redundant. This is inline with the novelty-based definitions of creativity and prior research [9, 26].

We can visualize these three qualities (creativity, usefulness, feasibility) as a spider plot as shown in Figure 10. Ideas can be mapped as a shape inside the triangle depending on their scores in each of these three areas. An idea that is truly innovative would have high scores in all three areas.

**Figure 10. A MAP OF PRODUCT INNOVATION AS A SPIDER PLOT**

An idea can be deemed creative without having a practical application, just as an idea can be deemed creative without being feasible. The authors define an innovative product idea as one that is creative/novel, feasible and of practical use as shown in Figure 11. Ideas that are creative and feasible, but not useful can go by several names including: novelty items, aesthetic innovation [27], or chindogu [28]. These are depicted in Figure 12. Ideas that are creative and useful but not feasible are desired but less attainable. We call these ideas chimera or pipe dreams and they are depicted in Figure 13. Ideas that are useful and feasible, but not novel are most likely existing products. This is depicted in Figure 14. It is possible to have ideas that score high on only one of the three metrics. The most common are ideas that are novel but not feasible or useful. These ideas are typically nonsense.

**Figure 11. EXAMPLE OF INNOVATIVE PRODUCT IDEA SCORE AND INNOVATIVE IDEA SKETCH**

**Figure 12. EXAMPLE OF NOVELTY ITEM IDEA SCORE AND NOVELTY ITEM IDEA SKETCH**

**Figure 13. EXAMPLE OF CHIMERA IDEA SCORE AND CHIMERA IDEA SKETCH**

**Figure 14. EXAMPLE OF EXISTING PRODUCT IDEA SCORE AND EXISTING PRODUCT IDEA SKETCH**
CONCLUSIONS
In this study, we have supported the notion that prolific idea generation is creative idea generation, as the quantity of ideas generated by the individual subjects had a very strong correlation with that subject’s overall creativity scores (r²=.82). This study also supports the novelty-based definition of creativity as reviewers’ subjective ratings of idea creativity had a strong correlation with ratings of idea novelty (r²=.80), but negligible correlation with idea usefulness (r²=.16). We found that laymen reviewers were not able to accurately rate ideas for feasibility or marketability, and we suggest experts (engineers or designers) rate such categories. We found that sketch clarity affected the perceived creativity. This supports an argument that engineers should have basic drawing abilities so their creative ideas are not overlooked. Finally, we suggest three independent qualities that fully describe an innovative product idea: creative (as a subjective judgment), useful (as defined as having practical applications), and feasible (as determined by experts).

We found that online reviewers are able to rate between 7-10 ideas per minute using five metrics. This is a rapid means of taking a large collection of ideas and reducing them to a manageable set of the most promising ideas.

REFERENCES
[1] Dean, D. L., Hender, J. M., Rodgers, T. L. and Santanen, E. L. Identifying quality, novel, and creative ideas: Constructs and scales for idea evaluation. Journal of the Association for Information Systems, 7, 10 2006), 646-698.
[2] Amabile, T. Creativity in context. Westview Press, Boulder, Colo., 1996.
[3] Amabile, T. M. Social-Psychology of Creativity - A Consensual Assessment Technique. Journal of Personality and Social Psychology, 43, 5 1982), 997-1013.
[4] Besemer, S. Creating Products in the Age of Design: How to Improve Your New Product Ideas! New Forums Press, Stillwater, OK, 2006.
[5] Besemer, S. P. Creative product analysis matrix: Testing the model structure and a comparison among products - Three novel chairs. Creativity Research Journal, 11, 4 1998), 333-346.
[6] Besemer, S. P. and O’Quin, K. Confirming the three-factor creative product analysis matrix model in an American sample. Creativity Research Journal, 12, 4 1999), 287-296.
[7] Besemer, S. P. and Treffinger, D. J. Analysis of Creative Products - Review and Synthesis. Journal of Creative Behavior, 15, 3 1981), 158-178.
[8] Besemer, S. and Oquin, K. Analyzing Creative Products - Refinement and Test of a Judging Instrument. Journal of Creative Behavior, 20, 2 1986), 115-126.
[9] Christiaans, H. Creativity as a design criterion. Creativity Research Journal, 14, 1 2002), 4154.
[10] Cropley, A. and Cropley, D. Engineering Creativity: A Systems Concept of Functional Creativity. Lawrence Erlbaum, City, 2005.
[11] Horn, D. and Salvendy, G. Consumer-based assessment of product creativity: A review and reappraisal. Human Factors and Ergonomics in Manufacturing, 16, 2 2006), 155-175.
[12] Horn, D. and Salvendy, G. Measuring Consumer Perception of Product Creativity: Impact on Satisfaction and Purchasability. Human Factors and Ergonomics in Manufacturing, 19, 3 2009), 223-240.
[13] Shah, J. J. and Vargas-Hernandez, N. Metrics for Measuring Ideation Effectiveness. Design Studies, 24,2003), 111-134.
[14] Shah, J. J., Kulkarni, S. V. and Vargas-Hernandez, N. Evaluation of idea generation methods for conceptual design: Effectiveness metrics and design of experiments. Journal of Mechanical Design, 122, 4 2000), 377-384.
[15] Ward, W. C. and Cox, P. W. Field Study of Nonverbal Creativity. Journal of Personality, 42, 2 1974), 202-219.
[16] Butler, D. L. and Kline, M. A. Good versus creative solutions: A comparison of brainstorming, hierarchical, and perspective-changing heuristics. Creativity Research Journal, 11, 4 1998), 325-331.
[17] Simon, H. A. Structure of Ill Structured Problems. Artif. Intell., 4, 3-4 1973), 181-201.
[18] Michalko, M. Thinkertoys : a handbook of creative-thinking techniques. Ten Speed Press, Berkeley, Calif., 2006.
[19] Ulrich, K. T. and Eppinger, S. D. Product design and development. McGraw-Hill Higher Education, Boston, 2008.
[20] Luecke, R. and Katz, R. Managing Creativity and Innovation. Harvard Business School Press, Boston, 2003.
[21] Nussbaum, B., Berner, R. and Brady, D. Get Creative! How to Build Innovative Companies. Bloomberg L. P., City, 2005.
[22] Osborn, A. F. Applied imagination; principles and procedures of creative problem-solving. Scribner, New York, 1963.
[23] Diehl, M. and Stroebe, W. Productivity Loss in Brainstorming Groups - Toward the Solution of a Riddle. Journal of Personality and Social Psychology, 53, 3 (Sep 1987), 497-509.
[24] Bray, R. M., Kerr, N. L. and Atkin, R. S. Effects of Group-Size, Problem Difficulty, and Sex on Group-Performance and Member Reactions. Journal of Personality and Social Psychology, 36, 11 (1978), 1224-1240.
[25] Mednick, S. A. The associative basis of the creative process. Psychol Rev, 69(May 1962), 220-232.
[26] Runco, M. A. and Charles, R. E. Judgments of Originality and Appropriateness as Predictors of Creativity. Personality and Individual Differences, 15, 5 (1993), 537-546.
[27] Eisenman, M. L. Essays on Aesthetic Innovation. Ph.D., Columbia University, NY, 2006.
[28] Patton, A. and Bannerot, R. Chindogu: A Problem Solving Strategy for Transforming Uselessness into Fearlessness. City, 2002.
[29] Rietzschel, E. F., Nijstad, B. A. and Stroebe, W. Productivity is not enough: A comparison of interactive and nominal brainstorming groups on idea generation and selection. Journal of Experimental Social Psychology, 42, 2 (2006), 244-251.
[30] O'Quin, K. and Besemer, S. Creative Products. Academic Press, City, 1999.
[31] Fink, A. and Neubauer, A. C. Neuroscientific Approaches to the Study of Creativity. City, 2006.
[32] Reing, B. A. and Briggs, R. O. Measuring the Quality of Ideation Technology and Techniques. City, 2006.
[33] Maccrimmon, K. R. and Wagner, C. Stimulating Ideas Through Creativity Software. Management Science, 40, 11 (1994), 1514-1532.
[34] Guilford, J. P. The Structure of Intellect. Psychological Bulletin, 53, 4 (1956), 267-293.
[35] Torrance, E. P. Predictive Validity of Torrance Tests of Creative Thinking. Journal of Creative Behavior, 6, 4 (1972), 236-252.