The Thesis Committee for Paul T. Williams
Certifies that this is the approved version of the following thesis:

Investigating the Innovation Capabilities of Undergraduate Engineering Students

APPROVED BY
SUPERVISING COMMITTEE:

Supervisor:

Carolyn C. Seepersad

Richard H. Crawford

Investigating the Innovation Capabilities of

Undergraduate Engineering Students

by

Paul T. Williams, B.S.M.E.

Thesis

Presented to the Faculty of the Graduate School of

The University of Texas at Austin

in Partial Fulfillment

of the Requirements

for the Degree of

Master of Science in Engineering

The University of Texas at Austin

August 2013

Dedication

To my family, for their constant support and encouragement

Acknowledgements

I would like to gratefully acknowledge my graduate advisor, Dr. Carolyn Conner Seepersad, whose guidance throughout my research has been indispensable. I am also thankful for Dr. Richard Crawford's instruction and support during my academic career. I would like to thank our collaborators at the University of Massachusetts Dartmouth, Dr. Trina Kershaw and Adam Young, whose time and input on the project offered new insights. Finally, I would like to thank the National Science Foundation for supporting the grant that funded this research.

Abstract

Investigating the Innovation Capabilities of Undergraduate Engineering Students

Paul T. Williams, M.S.E.

The University of Texas at Austin, 2013

Supervisor: Carolyn C. Seepersad

This thesis describes a method for measuring the innovation capabilities of mechanical engineering students and presents the results of a yearlong experiment. A review of relevant literature shows that it is unclear whether the innovation capabilities of engineering students increase or decrease over time. Experiments were conducted at two universities in which students were asked to redesign an everyday electromechanical product in a sketch-based concept generation activity. Student participants were also asked to complete a self-efficacy survey. Nearly one thousand concepts were generated from a combination of freshmen and seniors. The concepts were evaluated for originality, technical feasibility, and innovation characteristics by multiple raters. At both schools, the findings suggest that the senior-level engineering students are more creative than their freshman-level counterparts without sacrificing technical feasibility. Additionally, the seniors rated higher for originality at the end of the semester than they scored prior to taking their senior design class. These results suggest that the mechanical

engineering curricula, and especially the senior-level Engineering Design courses, are having a positive effect on student creativity.

Table of Contents

List of Tables

List of Figures

Chapter 1: Introduction

1.1 The Need for Innovation

There will always be a demand for innovative engineers as companies try to maintain a competitive edge in the global marketplace. As Dr. Charles Vest, President of the National Academy of Engineering stated, "In most areas of need, innovation is the only mechanism that can actually change things in substantive ways. Innovation is where creative thinking and practical know-how meet to do new things in new ways, and old things in new ways. The U.S. is still the most innovative nation on the planet. But we can only maintain that lead if we invest in the people, the research that enable it and produce a policy environment in which it can thrive [1]." During President Obama's recent visit to the National Academy of Sciences to celebrate its 150th anniversary, he highlighted America's "innovation that powers our economy and improves our health, protects our environment and security, [and] makes us the envy of the world [2]." In a previous visit to the Academy, he used the term "innovate" and its derivatives *twelve* times to refer to the pressing need to solve grand challenges in fields such as energy and health care [3].

In business, a failure to innovate and differentiate from competitors can cause products to become commodities [4]. To stay ahead of competitors, companies want to hire inventive and creative people for new product design and development. In an economic recession such as the one seen in 2009, collaborative team integration processes and new product development processes were required in order to be reasonably successful [5]. Curtis Carlson, chief executive of SRI International, put it clearly when saying, "Innovation is now the only path to growth, prosperity, environmental sustainability and national security for America. But it is also an incredibly competitive world. Many information industries require that products be improved by 100 percent every 12 to 36 months, just for the company to stay in business [1]."

This improvement in products can be either incremental or radical. Both radical and incremental product innovation "improve firms' performance because firms enhance their competitive advantage by differentiating new products from their competitors [6]." Concurrent to improvements on existing products, firms also want to introduce new products with superior benefits to increase their market share and financial performance. Along with successful new products, companies even strive to have the *reputation* of being innovative. "To many firms, a reputation as an innovative company is something that is both prized and actively sought after [7]." This common desire for an innovative reputation could be due to the fact that in one study, it was shown that product creativity to some degree influences the level of satisfaction with and willingness to purchase consumer products [8].

Cooper [9] likened the competitive market to warfare, with the senior executives acting as "generals" and newly developed products the "weapons" used to "invade" marketplaces. However, he says the battle is often decided by the "infantry": the engineers and scientists in R&D labs. This analogy might be extreme, but the fact remains that new products are the key to corporate prosperity. Clearly, innovation and creativity are important in both the corporate and educational environment.

1.2 Innovation Capabilities of Engineers Over Time

Creativity has often been studied as a static trait or as subject to change over long durations through training or education [10]. This study looks to measure changes in creativity over both long (multi-year) periods and also short (one semester) periods. Many factors can affect a person's creativity, such as work experience, maturity, and education. In a design exercise like the one performed for this experiment, previous exposure to many existing designs might give the engineer a larger collection of ideas to choose from. On the other hand, he or she may become fixated on a singular design and block out other potential solutions.

Other studies have been conducted in an attempt to quantify creativity at various stages of a student's academic career with mixed results. Cross *et al.* [11] compared second and final year engineering student performance on a design problem and found that the final year students produced more creative solutions. Atman *et al.* compared freshmen and seniors using a playground design problem [12]. The researchers observed student participants and then coded the process to see the design steps taken and time spent on each step, along with transitions between the steps. They found that senior performance was better than freshman performance with respect to several design elements, but there still remained room for improvement. In the same study, self-surveys did not correlate with solution quality, implying students were not very successful at evaluating their own abilities. When Lammi and Gero [13] compared design cognition between high school seniors and university sophomores, they found that the students shared commonalities in design but the university students spent more effort on analysis while the high school students expended more effort on synthesis. This over-emphasis on detailed analysis by the older students could become an obstacle if the goal is to brainstorm as many solutions as possible in a given amount of time.

The Accreditation Board for Engineering and Technology (ABET) requires that students meet certain criteria, stated as Student Outcomes, by the time of graduation [14]. For this study, the most relevant of these Outcomes are "an ability to design a system, component, or process to meet desired needs within realistic constraints" and "an ability to identify, formulate, and solve engineering problems." The requirements go on to state that the instruction must include engineering design topics wherein "engineering design is the process of devising a system, component, or process to meet desired needs. It is a decision-making process (often iterative), in which the basic sciences, mathematics, and the engineering sciences are applied to convert resources optimally to meet these stated needs."

To help students solve open-ended engineering problems, engineering curricula often include multiple design courses. In these engineering design courses, idea generation is an important part of the instruction. However, there is still a widespread feeling that the intellectual content of design is consistently underestimated. Project-based learning and design thinking are beginning to be introduced earlier in the engineering curriculum, with schools such as Penn State University and Northwestern

University requiring them in the first year [15]. A goal of this study is to quantify how well these design-based engineering courses are working.

1.3 How Creativity Is Measured

Many studies have been conducted on easily quantifiable measures, such as verbal and quantitative capabilities on standardized tests throughout elementary and high school, entrance exams for college or graduate school, and professional license exams. However, there is no universal standard for measuring creativity as it can manifest in many forms. As Zeng *et al*. suggested, creativity is a multifaceted phenomenon, which renders it complex and suggests that a definition of creativity should depend on specific research interests [16]. For mechanical design, Shah *et al*. identified a number of cognitive skills relevant to conceptual design [17]. They are now in the process of developing tests to evaluate these skills, beginning with Divergent Thinking [18].

In this study, creativity is documented with an Originality rating for each idea generated by the student participants. However, originality is not the only predictor of commercial success; many new products fail [9]. Therefore, new ideas must be technically feasible and also marketable to succeed in a competitive environment. This study aims to account for this aspect of design by also measuring the technical feasibility of student-generated concepts.

The motivation for the following study came from similar research completed at The University of Massachusetts Dartmouth (UMD) in 2011. Genco *et al*. [19] collected data that "suggest[ed] that freshman engineering students can be more innovative than their senior-level counterparts." This result was somewhat surprising and warranted further investigation. The researchers wished to see if this trend would be duplicated at another institution, namely the University of Texas at Austin (UT) or whether it was unique to the UMD group studied. The two universities differ in many aspects, such as overall size, research focus, and student body demographics. Furthermore, the increased pool of potential participants at UT was expected to produce a larger, more reliable sample size for the experiments.

As previously stated, innovation is critical to economic survival. This essential innovation will not be expected to come only from the experienced senior members of an engineering firm: new hires will be expected to contribute as well. Therefore, engineering graduates need to be prepared to generate creative ways to solve open-ended problems they will be presented with when entering their chosen industry. Accordingly, engineering departments need to be able to assess how well they are preparing students to become creative problem solvers.

This research proposes to define a process for measuring creativity at various stages of undergraduate engineering students' careers and use that data to check for growth in innovation capabilities. Simultaneously, the data can be used to measure the effectiveness of individual design instruction courses. The long-term goal of the project is to improve the curricula of those design courses to better develop students' creativity.

1.4 Research Statement

This study aims to measure student innovation in order to compare freshmen and seniors at two universities: the University of Texas at Austin and the University of Massachusetts Dartmouth. The results of this experiment will be used to evaluate the impact of the undergraduate engineering curriculum on the innovation capabilities of students.

To formalize this goal, the researchers developed the following research statement:

> *By measuring the innovation capabilities of undergraduate engineering students at different stages of their undergraduate education, we can determine whether those capabilities are affected by undergraduate educational experiences.*

> Furthermore, the underlying hypothesis to be tested can be stated as:

> *The innovation capabilities of final-year undergraduate mechanical engineering students are greater than those of first-year undergraduate mechanical engineering students.*

To test this hypothesis, groups of students were asked to participate in a concept generation exercise. This experiment was a sketch-based idea generation activity using an everyday household appliance. The concepts produced in the experiments were then analyzed for originality, technical feasibility, and innovation characteristics to detect trends between various groups of students.

A literature review of related material will be conducted in Chapter 2. In Chapter 3, a detailed account of the research method will be presented. Chapter 4 will include the data and results from the conducted experiments, while Chapter 5 will summarize the conclusions drawn and provide suggestions for future work.

Chapter 2: Literature Review

Research relevant to the current investigation has been conducted in the area of creativity measurement and evaluation. The following sections provide an overview of selected topics related to the research method to be described in Chapter 3. First, factors that may affect creativity, such as design fixation, are explored. Then, known tools for increasing creativity are reviewed before concluding with various methods of measuring creativity.

2.1 Factors That May Affect Creativity

Many factors can affect a person's creativity. The goal of an engineering design course is often to introduce divergent thinking and encourage the student to consider ideas that are not obvious [15].

A major barrier to creativity, especially in concept generation, can be design fixation. Indeed, Condoor *et al.* found that engineering students are often hampered by design fixation [20]. Fixation has been qualitatively observed to limit creative output during the conceptual design process when designers unintentionally adhere to a constrained set of ideas [21]. In an experiment conducted by Purcell *et al.*, design fixation was found in mechanical engineers but not industrial designers [22]. However, this might have been because the industrial designers in the experiment were fixated on "being innovative" and creating as many different ideas as possible rather than creating a well-developed solution like the engineers were. Gero adapted linkographs in an attempt to quantify design fixation, although his system is still in an early stage and more research is needed [23]. Encouragingly, some of the problems caused by design fixation can be mitigated (although not eliminated) with methods of "defixation" [24]. Chrysikou and Weisberg found that if a group of students was shown a problematic example design accompanied by a description of its elements (some of which are problematic), many of the problematic elements were still present in the students' solutions, due to fixation. However, if the description was accompanied by instructions specifically to avoid using its problematic elements, the students became "defixated" and the quality of their solutions improved.

Traditional thinking assumes that engineering students acquire cognitive skills, such as creativity, as they gain knowledge and design skills [25]. However, Lai *et al.* [26] found that older students are trained to think in a certain way and therefore freshman students may be better equipped to solve open-ended problems using creative thinking. As Guilford [27] found, acquiring skills and knowledge does not necessarily increase creativity. He found a curvilinear relationship between intelligence and creativity: people with low levels of intelligence were unlikely to be creative but the creative output of those at high intelligence levels varied from low to high levels of creativity.

Green *et al.* [10] studied whether explicit cues to "think creatively" can improve short-term creativity. They found that cues were successful in eliciting responses that were not only novel but were also appropriate to the task constraints.

In general, these studies show that mechanical engineers are not necessarily creative by nature. During creative activities such as concept generation, they may need to overcome a tendency to fixate on an initial and often suboptimal solution. These studies also imply that it is unclear how the creative capabilities of engineers change with time. Some studies suggest that curricula can have a tendency to "funnel" students towards a certain way of thinking. Others show that creative capabilities can improve with exposure to existing designs and idea-generation techniques. This study attempts to measure the creativity of mechanical engineering students over their academic careers to see if there is a significant change with time.

2.2 Known Tools for Increasing Creativity

There are various tools that have been shown to increase creativity when employed in the concept generation phase. They can be divided into many categories, such as intuitive and directed, or individual and group techniques. Example methods of stimulating creativity include 6-3-5, Design by Analogy, and Empathic Experience Design (EED). Figure 2.1 catalogues some of the many options available to design engineers.

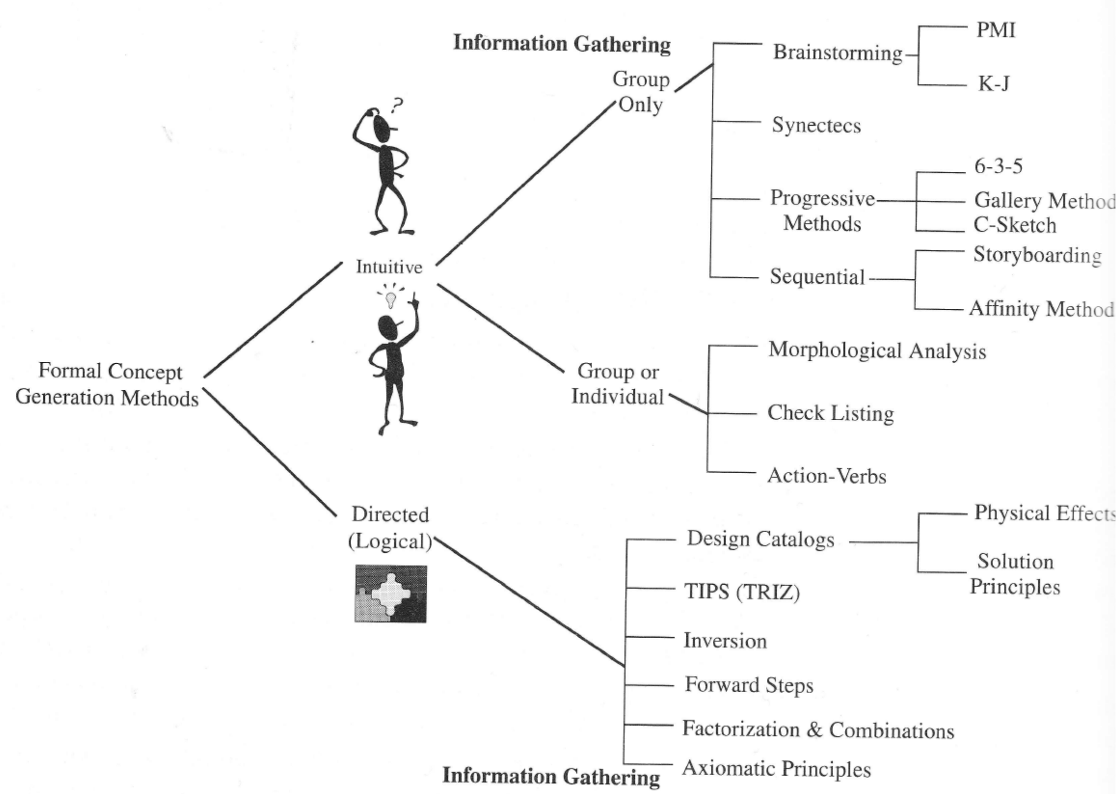

Figure 2.1: Concept Generation Techniques (from Otto & Wood [28])

The technique most relevant to this study is the 6-3-5 method [29], where 6 people individually generate 3 ideas and record them as phrases on a sheet of paper, then circulate those ideas to the other designers for 5 rotations. C-Sketch is another group technique very similar to 6-3-5 but each person only generates one concept in the form of a sketch before rotating [30]. An iterative process that essentially combines these two techniques, as suggested by Otto and Wood [28], is used in this study. There are further variations on these basic methods, such as when Wodehouse and Ion [31] looked to augment the 6-3-5 method with an ICR ("Inform, Create, Reflect") Grid. However, this is beyond the level of sophistication needed for the analysis described in Chapter 3.

Design by analogy techniques [32], [33], [34] attempt to inspire new ideas by bridging engineering domains. These analogies are also used for process planning, cost estimation, and evaluation of a new product. A growing subset of analogies can be found in bio-inspired or biomimetic designs, where designs found in nature are used to develop engineering solutions to new problems [35].

Empathic Experience Design is a structured conceptual design method that inspires designers by using empathic experiences to identify and address latent needs that customers cannot articulate [36]. Possible methods to simulate situational disabilities while using a product include the use of thick gloves or blindfolds. To identify and create a list of EED Innovation Characteristics (later used in this analysis), Saunders *et al.* studied nearly 200 innovative products [37]. These characteristics were organized into categories of functionality, architecture, external interactions, user interactions and cost. Of the 197 award-winning products evaluated in the study, approximately 75% exhibited at least three different characteristics of innovation and 95% exhibited at least two. The majority of these characteristics were demonstrated as external interactions and user interactions. From these results, the authors suggest that engineers may be wise to broaden their focus to multiple characteristics when attempting to design innovative products. The results also stress the need for engineering design methodologies that focus on improving product interactions, such as the aforementioned Empathic Experience Design.

Perhaps the most well-known form of concept generation is brainstorming. Brainstorming involves team members communicating ideas verbally during a set time period [28]. The goal of brainstorming is to obtain several concepts that might work as solution principles to a piece of the design problem. A variation of brainstorming is brainsketching. Brainsketching is an idea generation technique that uses sketching as the primary means of recording ideas [38]. Van der Lugt [39] found that it is not necessarily better than brainstorming but rather should be considered a different process that may serve a different purpose. Namely, it tends to offer more incremental improvements rather than drastic advancements. Methods of organizing brainstormed ideas for evaluation include mind maps [40] or a gallery [38].

The Theory of Inventive Problem Solving (abbreviated as TIPS, or TRIZ from the direct Russian translation) was developed as a formal method to inspire new ideas from conflicting design variables [41]. Thousands of patents were studied to observe historical trends in invention, which led to design principles that can be used to solve engineering conflicts. However, the effectiveness of TIPS can be limited by the designer's knowledge since the design principles often are vaguely worded in order to apply to many engineering domains.

Universities have adapted their curricula to incorporate these tools with the goal of enhancing students' creativities. Hirsch *et al.* [42] described an innovative, interdisciplinary, project-based freshman course at Northwestern University named Engineering Design and Communication (EDC). Design and communication are integrated from the beginning of the course and students are given a single letter grade combining both aspects of the course. The hope is that exposure to these various techniques early in a student's academic career will have a positive effect on his or her creativity.

Some research has been done in an attempt to figure out which creativity techniques work best in which setting. Linsey *et al.* [43] found that a rotational viewing of sets of sketches (such as 6-3-5 or C-Sketch) produces more ideas than having all concepts displayed in a "gallery view" form, but a gallery view results in more high quality concepts. Suwa *et al.* found that regrouping sketches creates new ideas through the detection of unintended features [44]. This outcome could manifest itself in this experiment if the participants see a new idea as the concept travels around the group. Conversely, it could unintentionally affect results if the rater incorrectly sees something more innovative than the student intended. Yang [45] found statistically significant correlations between the quantity of brainstormed ideas and the design outcome but not always a correlation between the quantity of morphological alternatives and design outcome. Mullen *et al.* [46] found that the productivity of group brainstorming is lower than the combined productivity of individuals working alone.

Overall, these comparison studies tend to indicate that rotational group techniques like 6-3-5 or C-Sketch are best for generating large quantities of ideas, such as those needed for analysis in this experiment. Therefore, the researchers decided to use a modified 6-3-5 method [28] in the research procedure. There is also minimal training needed for participants for a technique like 6-3-5, especially when compared to a method such as TIPS, where the results will depend on the user's expertise. Finally, the modified

6-3-5 technique allows the researchers to evaluate both individual and small group trends in the data.

2.3 Measuring Creativity

One significant hurdle when designing this experiment was finding a consistent way to measure creativity. The scales the researchers used are described in further detail in Chapter 3. Much like how Zeng *et al*. [16] define two aspects of creative products as "novelty and appropriateness," this study looks at originality and feasibility of the ideas generated by the freshman and senior participants.

Shah *et al*. defined metrics [47] for measuring ideation effectiveness, identifying Novelty, Variety, Quantity, and Quality of sets of designs as the categories to analyze. In their definitions, Novelty is how new or unusual an idea is compared to what is expected. Variety is measured as the extent to which the generated ideas span the solution space. Quantity is simply the total number of ideas generated during the designated amount of time. Finally, Quality measures how feasible the set of ideas is as well as the ideas' relative ability to satisfy design requirements. Guidelines are given for each rating but they refrained from consolidating the four scores into an *overall* effectiveness measurement. Nelson *et al* proposed a single, new metric (using Shah *et al*.'s metrics) to evaluate the quality of design space exploration [48]. This single metric, called DSQ (quality of design space exploration), attempts to provide a single, quantifiable point of comparison for characterizing the conceptual design process between multiple designers.

In another study, Oman *et al.* [49] proposed using two methods for creativity concept evaluation during early design: the Comparative Creativity Assessment (CCA) and the Multi-Point Creativity Assessment (MPCA) methods. The CCA metric, created as an attempt to assist designers and engineers in assessing the creativity of their designs quickly from the concept design phase, is a weighted combination of Shah *et al.*'s Novelty and Quality metrics described above. The MPCA metric is an adaptation of NASA's Task Load Index but instead of assessing workloads it indicates different aspects of creativity. It is a method for judges to break down concepts into seven aspects of creativity (original/unoriginal, well-made/crude, surprising/expected, ordered/disordered, astonishing/common, unique/ordinary, and logical/illogical) that factors in the judges' preferences of importance for those aspects. In the end, the MPCA method provides an easily understood creativity value calculated on a 0-10 scale, with 10 being the most creative.

Sarkar and Chakrabarti [50] chose to analyze novelty and usefulness of products to determine the overall creativity of products. They developed multi-step formulas to scale the relative Novelty and Usefulness of a group of products. Those two values are then multiplied to generate a Creativity score. Novelty is assessed using a qualitative scale of "very high," "high," "medium," or "low." Usefulness is defined as the product of level of importance, rate of popularity of use, frequency of usage, and duration of use. This technique appears best for determining relative creativity amongst a given set of designs but necessitates extensive training of the rater.

Charyton *et al*. [51], [52] developed instruments to measure the fluency, flexibility, and originality of creative outcomes. Their originality instrument provided the foundation for the originality metric used in this study. Their originality metric implemented an 11-level scale as shown in Table 2.1.

Table 2.1: Originality Scale from Charyton *et al*. [51]

Score	Description
0	Dull
1	Common
2	Somewhat interesting
3	Interesting
4	Very interesting
5	Unique and different
6	Insightful
7	Exceptional
8	Valuable to the field
9	Innovative
10	Genius

However, this scale is difficult to use because the descriptions do not have an intuitive order with respect to innovation levels. For example, it is not immediately clear why "Exceptional" is three points greater than "Very interesting" yet two points less than "Innovative." Additionally, previous research has shown that high fidelity scores lead to lower levels of repeatability [53]. Therefore, a 5-level scale was implemented in this study, as described in Section 3.5.2.

Creativity can also be measured indirectly through surveys, questionnaires, interviews, or other similar methods of gathering data from participants. There is some evidence that self-efficacy, or belief in one's abilities toward a given task, is a reliable predictor of design outcomes since confidence levels may limit a designer's scope or creativity [54]. However, in other experiments student self-surveys [12] and reflective essays [55] correlated poorly with actual performance. To determine if the participants of this experiment were able to accurately assess their own abilities, a student self-survey was included in the procedure.

Clearly, there is no universal standard for measuring creativity in engineering design. The selection of an appropriate metric is essential to each study and results may vary based on the system employed by the judges or evaluators. The following chapter includes a discussion of the creativity metrics used by the researchers in this study.

Chapter 3: Research Method

With the goal of comparing freshmen to seniors, student volunteers were solicited from general first-year and final-year mechanical engineering courses to participate in a design experiment. The experiment itself comprised a small-group concept generation exercise and a self-survey. After all groups completed the experiment, the data was evaluated for Originality, Feasibility, and Innovation Categories. In the previous research completed at UMD [19], the participant group consisted of 48 freshmen and 46 seniors. Those groups were then subdivided into "innovation enhancement" and "non-enhancement" groups, each consisting of 18 to 28 students. At UT (and in the most recent UMD experiments), no "innovation enhancement" techniques were used so the resulting data sets are larger. In total, 188 students from UT and 137 students from UMD participated in the exercise during the 2012-2013 academic year. The following sections focus on the details of the research method as it was conducted at UT, but the procedure at UMD was very similar.

3.1 Problem Definition

The design problem was focused on redesigning an everyday electro-mechanical product by small groups of students. The product used in the experiment needed to be familiar to a wide range of participants, inexpensive, easy to procure, and not overly complex. Since students would have only five minutes to interact with the product before beginning the concept generation phase, a completely unknown or overly complicated product would be a poor choice. In past Empathic Experience Design (EED) studies

[56], an alarm clock was one of a few products selected to meet these criteria. Therefore, the researchers decided to continue using alarm clocks as the product for this phase of research. There was no risk of some students having an unfair advantage by knowing the details of the experiment since all past participants were seniors at the time of experimentation and had since graduated. Another benefit of using clocks is that as college students, the participants all utilize some type of alarm system to wake up for classes and thus have ideas of what they like or would want to improve.

Figure 3.1: Current Generation Alarm Clocks

3.2 Participant Selection

At UT, volunteers for the experiment were solicited from ME302, Introduction to Engineering Design & Graphics, a course taken in either the first or second semester. The rationale for choosing this class is that these students have not yet been exposed to the other elements of the mechanical engineering curriculum. Potential volunteers were told that they would be asked to redesign a product in small groups. To prevent bias, no further information on the type of product was given. In exchange for participation, students would be provided with a free barbeque dinner (or a vegetarian option if desired).

Similarly, seniors were recruited from ME366J, Mechanical Engineering Design Methodology. This class was chosen because it is typically taken in the final year of mechanical engineering studies and thus the students have progressed through the mechanical engineering curriculum. In exchange for participating, the seniors were given individual extra credit on one of their group project deliverables. As an alternate extra credit option to the experiment participation, students could submit an essay of approximately 500 words. In this essay they had to identify three mechanical products that failed and discuss why they failed. This option was provided in case students had conflicts with every possible time slot or did not feel comfortable participating in the experiment and to avoid the appearance of coercion or undue influence to participate.

In the end, 54 students participated from the freshman level class and 58 from the senior level class for a total of 112 participants during the fall semester.

3.3 Differences between Fall and Spring

The goal stated in the research proposal was to get 50% of the volunteers from the beginning of the semester and 50% from the end of the semester so that the effect of the individual courses could be investigated as well. However, for the spring ("end of semester”) group, it was much more difficult to get freshmen to volunteer. The recruitment procedure was unchanged from the fall semester. The only change in the incentive was that the free meal was pizza instead of barbeque due to budget concerns. The lower numbers were possibly due to students being busier at the end of the semester with group project deadlines and finals approaching. They did not have as much desire to add an optional activity to their busy schedules. Therefore the researchers were only able to get 11 freshmen participants, resulting in 33 concepts. This was not a major concern since the freshman-level course does not have a significant design focus and the effect of completing the course on student creativity was expected to be minimal. The seniors, on the other hand, were still very interested in getting individual extra credit for their class. Another factor that contributed to the high senior participation rate is that the experiments were conducted during normal lecture time. The researchers were able to take advantage of an open class date due to the professor’s travel schedule. Thus, students definitely would not have an extracurricular conflict during that time and 65 seniors participated. Figure 3.2 shows the number of student volunteers per semester at UT.

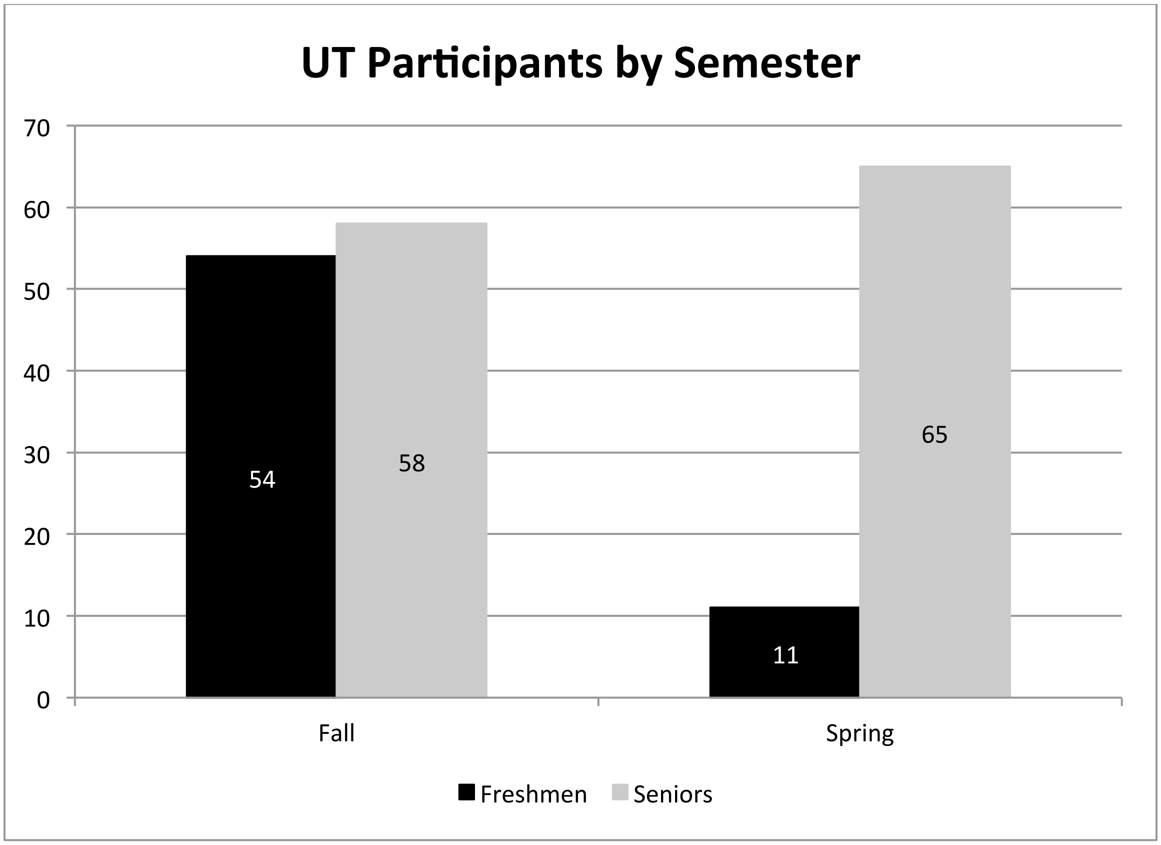

Figure 3.2: University of Texas Participants

3.4 Experiment

At each experiment session, the students were separated into small groups. Ideally these groups consisted of six students, but often there would be an uneven total due to cancellations. The small groups usually consisted of either five or six students, but there were a few rare occasions where it was necessary to have four or seven students in a group. With the freshman students, there tended to be a lot of cancellations on the day of the experiment, possibly since there was no significant penalty for missing their assigned time slot.

The students were each assigned a three-digit identification number. This number was used to keep the concepts confidential and reduce possible bias from having a rater recognize a student's name during evaluation. To comply with IRB protocol, students were handed a consent form as they entered the experiment to confirm that they were willingly participating in the study.

Once the students were in their groups, they were presented with the task to redesign an alarm clock. The same problem statement was read to all participants:

> *Thank you for helping with our study. We are trying to design the next generation of breakthrough alarm clocks and request your help in generating ideas. We have provided a standard alarm clock for you to interact with to get a better understanding for potential improvements and customer needs. We ask that each person interact with the alarm clock over the next five minutes and rotate the clocks around the group so that everyone gets a chance to interact with the alarm clock. Then, we will begin the design exercise. Before we begin are there any questions?*

After the prompt was read, the "current generation" alarm clocks were distributed and the students had five minutes to interact with the clocks before beginning the design exercise. The students' behavior varied during this time period. Some groups sat silently and allowed each student to individually interact with the clock. Other groups were very talkative and discussed features they liked or disliked on the clock they were holding or on their own clocks at home. Some groups walked to a wall outlet to plug in the clock to set the time, hear the alarm, and explore other functions of the clock.

While the students were interacting with the alarm clocks, blank sheets of paper and boxes of markers were distributed to the groups. Once the five-minute period of clock interaction was completed, the modified 6-3-5 exercise was explained to the students. The students were given 15 minutes to sketch 3 different concepts in their unique marker color on 3 separate sheets of paper. They were also asked to write their three-digit identification number in the corner of their sketches for tracking purposes. They were allowed to use phrases or comments to help convey their ideas, but it was emphasized that the ideas should mainly be described in the sketches.

At the end of this first round of 15 minutes, each student then passed the 3 concepts to the next student in the group. That student then made comments or sketches (again in his or her unique marker color) to expand, critique, or improve the previous student's ideas. The participants had five minutes for each round after the first round.

The concept generation phase concluded once each student had had the opportunity to expand on all other group members' ideas. The final step of the procedure was to have each student fill out a survey assessing his or her own creativity. Again, the three-digit number was used for identification when completing the survey. Partway through the fall semester, a new version of the survey was approved which included the participant's official UT rank (Freshman, Sophomore, Junior, or Senior) along with his or her GPA.

After the survey was completed, all concepts and surveys were collected for future analysis. The students were thanked for their participation and read the following statement:

We would like to thank you for participating and ask that you please refrain from discussing details of this experiment or the concepts your group generated with your classmates who have yet to participate. We also ask that you not discuss this experiment with future students as we may be doing similar experiments in the future.

After all experiments were finished, the concepts were organized into numeric order according to participant number and "–A" "–B" or "–C" was added to uniquely identify each concept. To aid in analysis, the paper copies were scanned to color PDF's and saved electronically.

3.5 Analysis Steps

3.5.1 Feature Definitions

For the alarm clocks, the concepts were divided into feature-level ratings, shown in Table 3.1, as used in previous studies at UT and UMD. This categorization helped the evaluators to consider all of the different aspects of the generated concepts. These feature definitions were chosen because previous research indicated that the metrics described in Section 3.5.2 were most effective when applied at the feature level [57] [53].

Table 3.1: Alarm Clock Feature Definitions (Adapted from Saunders [57])

Mode of Alarm: Any feature that is used to wake up the user. This also applies to any feature that forces the user to get out of bed, including moving alarm clocks. Standard feature defined as the use of beeping or ringing sound.
Display Type: Any form of displaying information to the user. Standard feature defined as basic analog, LED, or digital display.
Information Shown: The type and amount of information displayed to the user. Standard feature considered current time, alarm time (analog clocks), am/pm, and alarm on/off.
User Input: Any method of input or interaction between the user and clock, including setting the clock or turning the alarm on/off. Standard features include buttons or dials.
Energy source: Any feature that supplies power to the clock. Standard features defined as power cord or battery.
Snooze: Any feature that describes whether an alarm clock has snooze capabilities and how to activate the function. Standard features include buttons for digital clocks and none for analog clocks.
Music Player: Any feature that describes the existence or type of music playing capabilities. Standard feature implies no music playing capabilities. Music is separate from additional uses because of the prevalence of music playing functionality.
Additional Uses: Describes any additional functions performed by the alarm clock beyond traditional and music playing capabilities. Standard feature implies no additional functions.
Shape/layout: Describes the architecture, outward appearance, or theme of the clock frame. Standard implies no unique characteristics of the shape or layout.

3.5.2 Metrics

The next step in the analysis was to find a set of metrics with which to rate the concepts for originality and feasibility. In previous research, Genco, Hölttä-Otto and Seepersad had success utilizing a 5-level scale for originality and a 4-level scale for feasibility [19]. The originality scale is similar to commonly applied five-point Likert [58] scales, one of the most recommended scales for use in rating systems [59]. It is also very similar to a novelty metric proposed by Shah *et al*. [47], designed to compare each concept to a set of predefined solutions or to other concepts generated by the same experiment. The interval definitions for the originality scale are shown in Table 3.2.

Table 3.2: Originality Scale

Originality Scale	
0	Common
2.5	Somewhat Interesting
5	Interesting
7.5	Very Interesting
10	Exceptional

As the table shows, exceptional concepts were given a rating of 10 while the concepts that showed common or standard existing features were given no credit and a rating of 0. As an effort to increase consistency and prevent bias while rating hundreds of concepts over a yearlong study, a list of exemplars for the clock features listed in Table 3.1 was created. The list was helpful when a rater was unsure of what rating to apply to a current concept; he or she could see which previously rated concept it most resembled in terms of originality. The exemplar list, found in Appendix A, was initially created by reviewing concepts generated in past EED experiments. These concepts had been rated using the same system by previous collaborators and provided a basis to determine whether a concept deserved non-zero originality points. A list of standard features that earn zero originality points, some of which are described in Table 3.1, was created by studying existing alarm clocks from sellers such as Target, Wal-Mart, and

Amazon.com. Sample evaluations for two concepts are described in detail in Section 3.5.2.1.

As another advantage, this exemplar list of originality ratings can be used as a guideline or reference point for future raters. Furthermore, it proved helpful when used to train new raters without engineering backgrounds. However, it is not intended to be an absolute set of rules. For example, if the mode of alarm of “vibrate bed” was categorized in the exemplar list as a rating of 5 but was done in a unique or well-developed way, the concept might merit a 7.5 or even 10 rating when evaluating the overall concept.

The technical feasibility was assessed based upon the method suggested by Shah [47] and modified by Linsey [32]. A flow chart was used to determine the feasibility rating (shown in the blocks on the right of Figure 3.3) that each concept should garner.

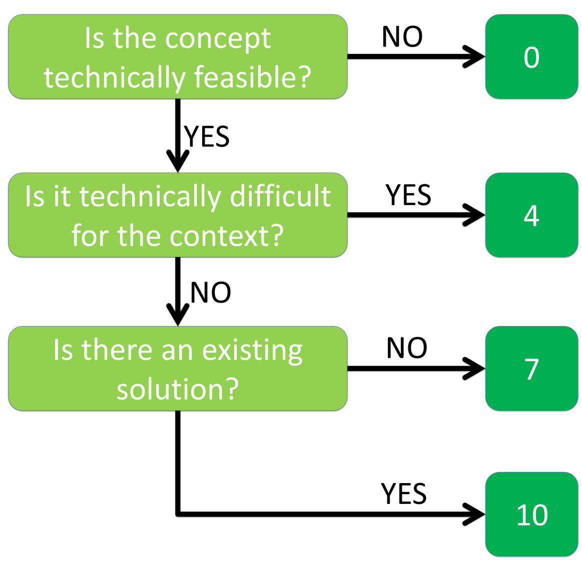

Figure 3.3: Feasibility Scale

As the chart shows, the ideal score is again 10 out of 10. In this analysis, concepts were considered feasible if similar features are present in the consumer market, even if not currently in alarm clocks. For example, if the student drew a concept that involved a tilting bed, a manufacturer could adapt that technology from a hospital bed to an "alarm clock bed" without unreasonable difficulty.

Each concept was rated on a feature-level basis and then given an overall Originality and Feasibility rating. The overall Originality score was taken as the maximum of all the feature originality scores and the overall Feasibility score was taken as the minimum of the feature feasibility scores. The reasoning behind this scoring system was that the overall scores should reflect the most innovative feature while also accounting for the least technically feasible aspect of the design.

As a third metric, Innovation Categories were assigned to each concept that rated above zero in Originality. These Innovation Categories were established by Saunders *et al*. [37] and are presented in Table 3.3. While they were not the main focus of this experiment, they can still provide an interesting measurement of creativity. If a characteristic (shown as a bullet point in Table 3.3) was present in the student's concept, it was marked with a "1" in the corresponding column of the score sheet. Some of the Innovation Categories are composed of multiple Innovation Characteristics. Concepts with three or more Innovation Categories were considered "high level" concepts in subsequent analysis.

Table 3.3: Innovation Categories & Characteristics (Adapted from Saunders *et al*. [37])

Functionality

- *Additional Function*: Allows the user to solve a new problem or perform a new function while still performing the function of the comparison product.

Architecture

- *Modified Size*: The physical dimensions during operation or storage have changed in expansion or compaction beyond subtle or incremental differences.
- *Modified Physical Layout*: The same elements of the product are still present, but the physical architecture has changed.
- *Expanded Usage Physical Environment*: The product can now be used in more usage environments with different resource availability or different physical characteristics.

External Interactions

- *Modified Material Flow*: Accepts or creates different materials or uses materials in new ways.
- *Modified Energy Flow*: Utilizes new sources of energy or converts to a different form of energy than previously used.
- *Modified Information Flow*: Different types or amounts of information are being gathered, processed, or output/displayed.
- *Interaction with Infrastructure*: The product interacts with previously owned infrastructure.

User Interactions

- *Modified Physical Demands*: The product is easier to use physically beyond subtle or incremental differences.
- *Modified Sensory Demands*: The product is easier to use from a sensory stand point beyond subtle or incremental differences.
- *Modified Cognitive Demands*: The product is easier to use mentally beyond subtle or incremental differences.

Cost

- *Purchase Cost*: Purchase cost is significantly different.
- *Operating Cost*: Operating and/or maintenance costs are significantly different.

3.5.2.1 Sample Analysis

Two concepts were selected to show how the preceding metrics are typically applied by the raters. Two next-generation alarm clock concepts from UT students are shown in Figures 3.4 and 3.5. Corresponding originality and feasibility ratings are shown in Table 3.4 and innovation characteristics are summarized in Table 3.5.

The first concept is an above-average concept. It earned an Originality score of 7.5 out of 10, a Feasibility score of 10 out of 10 and 3 Innovation Categories. The concept is significantly different from a typical alarm clock. The original idea, shown in red, is a bed that vibrates to shake the user awake. As the sketch was passed around the group, the idea was expanded to include a wristband that sends mild shocks to the user. The combination of these two features earned the high rating for "Mode of Alarm." While most of the other categories are either "standard" or not present in the sketch, the shape/layout of a mattress is also original when compared to standard alarm clocks. Since the overall Originality score takes the maximum of all features, the final rating is 7.5. The overall Feasibility score was 10 because no feature was deemed technically difficult for the context and present technology (such as a dog collar) could be adapted safely for personal use. This design exhibits five Innovation Characteristics: Modified Size (mattress-sized), Modified Energy Flow (electric energy is being converted to vibrations), Interaction with Infrastructure (interacts with an existing bed), Modified Physical Demands (the user must get out of bed), and Modified Sensory Demands (the user senses the vibrations and shocks). However, these 5 Innovation Characteristics only account for 3 Innovation Categories: Architecture, External Interactions, and User Interactions.

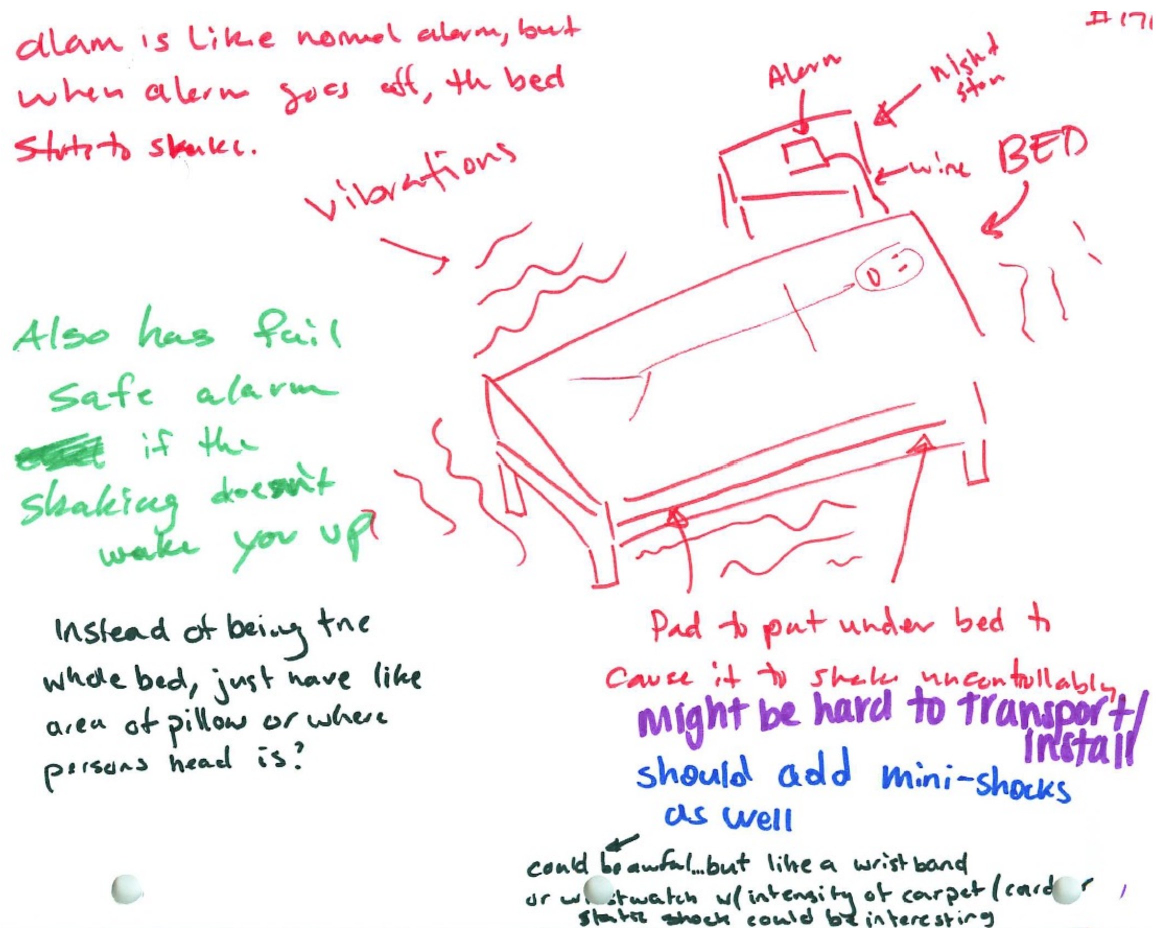

Figure 3.4: Sample Concept 1

The second concept is an average concept. It earned an Originality score of 2.5 out of 10 and a Feasibility score of 10 out of 10. Even though it only scored 2.5 in Originality, it still had 3 Innovation Categories. Most of its features can be found in standard alarm clocks and thus received zero originality points. Wheels were added to make the clock roll away from the user, which earned 2.5 originality points for being "somewhat interesting" although there have been clocks such as the "Clocky" [60] introduced with this feature. Additionally, having the clock be solar-powered earned 2.5 originality points in the "Energy Source" feature. The solar power feature exhibits two Innovation Characteristics: Modified Energy Flow (solar energy is the power source) and Expanded Environment (the clock could be used in non-traditional settings, such as a camp site). The rolling feature adds an additional Innovation Characteristic: Modified Physical Demands (the user must walk around to find the clock).

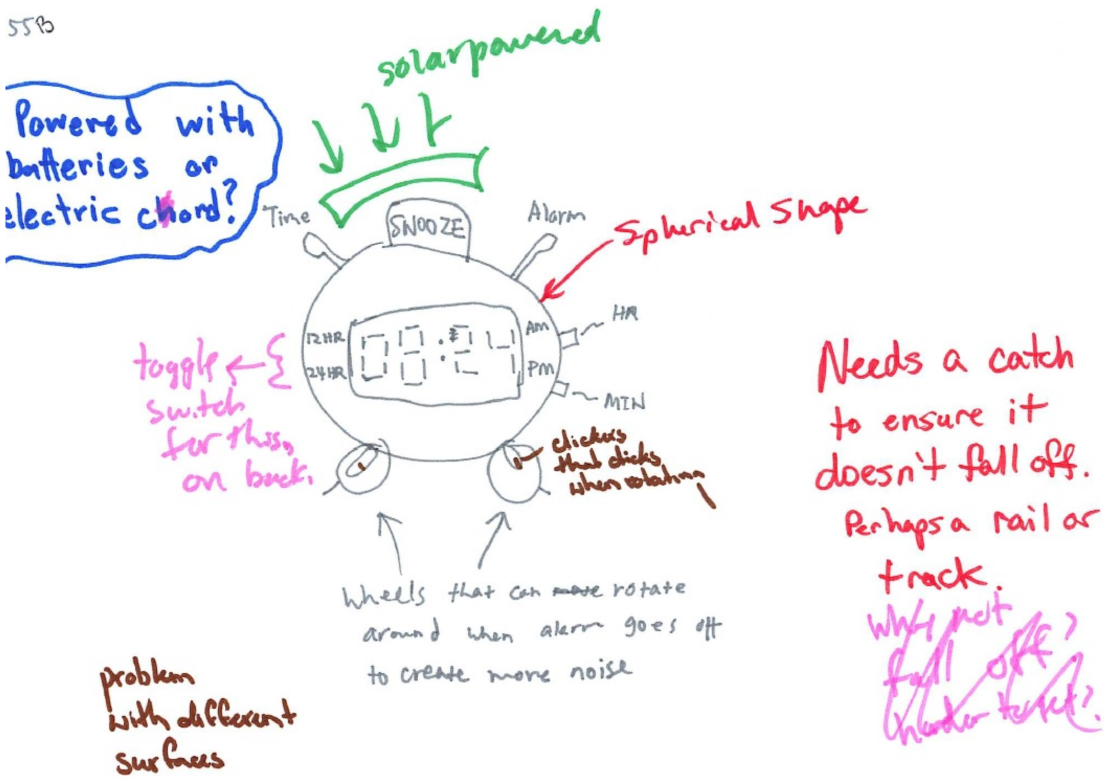

Figure 3.5: Sample Concept 2

Table 3.4: Sample Originality and Feasibility Ratings

Design	Category	Mode of alarm	Display type	Info shown	Mode of input	Snooze	Energy source	Music	Alterna-tive use	Shape/ layout
Concept 1	**Feature**	vibrate bed, wristband shock	standard	standard	standard	standard	standard	standard	standard	mattress
	Originality	7.5	0	0	0	0	0	0	0	5
	Feasibility	10	10	10	10	10	10	10	10	10
Concept 2	**Feature**	rolls	standard	standard	standard	standard	solar	standard	standard	sphere
	Originality	2.5	0	0	0	0	2.5	0	0	0
	Feasibility	10	10	10	10	10	10	10	10	10

Table 3.5: Sample Innovation Categories and Characteristics Ratings

Design	Functionality	Architecture			External Interactions				User Interactions			Cost
	Additional Function	Modified Size	Modified Layout	Expanded Environment	Modified Material Flow	Modified Energy Flow	Modified Information Flow	Interaction with Infrastructure	Modified Physical Demands	Modified Sensory Demands	Modified Cognitive Demands	Cost
Concept 1		1				1		1	1	1		
Concept 2				1		1			1			

3.5.3 Inter-Rater Reliability

To acquire more consistent ratings, each set of concepts was analyzed separately by two raters. The reliability between raters was evaluated using Cohen's weighted kappa [61], which ranges from 0, which represents chance agreement, to 1, which represents perfect agreement. The weighted kappa measurement was chosen because it accounts for chance agreement between raters and includes a provision for scaled disagreement. Namely, a disagreement in ratings of 0 versus 7.5 is penalized more than a disagreement of 5 versus 7.5. Fleiss [62] defined an inter-rater agreement of 0.6-0.8 as "substantial" while Landis and Koch [63] defined an agreement of 0.75 or higher as "excellent." Therefore, the researchers set a goal of reaching above 0.75 for inter-rater agreement, but values above 0.6 would be considered acceptable to begin the rating process. The calculations for the metric are shown in Equations 3.1 to 3.5, where κ_w is Cohen's weighted kappa coefficient, p_{oij} is the ratio of observed agreements for a given combination of ratings (i and j) of two raters, p_{cij} is the ratio of agreement due to chance between two raters for a given set of ratings (i and j), $P(Rater_x)$ is the probability that a given rater will assign any ranking, and v_{ij} is the weight of the disagreement between two raters for a particular set of ratings (i and j).

$$\kappa_w = 1 - \frac{q'_o}{q'_c} = 1 - \frac{\sum v_{ij} p_{oij}}{\sum v_{ij} p_{cij}} \tag{3.1}$$

$$p_{oij} = \left(\frac{number\ of\ observed\ agreements}{number\ of\ possible\ agreements}\right)_{ij} \tag{3.2}$$

$$p_{cij} = P(Rater_1)_i * P(Rater_2)_j \tag{3.3}$$

$$P(Rater_x) = \frac{number\ of\ assignments\ by\ Rater\ x}{number\ of\ possible\ assignments} \tag{3.4}$$

$$v_{ij} = (i - j)^2 \tag{3.5}$$

For the Innovation Category ratings, the standard kappa metric was used because each Category is a binary quantity and thus does not warrant the scaled disagreement factor. The calculations for the metric are shown in Equations 3.6 to 3.8 where κ is Cohen's kappa coefficient, p_o is the ratio of observed agreements, p_c is the chance agreement and $P(Rater_x)$ is the same as previously defined. Since empathic experience features were not the focus of this research, these inter-rater metrics were used while training the UT raters but were not expanded to the UMD raters. Therefore, the Innovation Category ratings found in Chapter 4 are from a single rater.

$$\kappa = \frac{p_o - p_c}{1 - p_c} \tag{3.6}$$

$$p_o = \frac{number\ of\ observed\ agreements}{number\ of\ possible\ agreements} \tag{3.7}$$

$$p_c = P(Rater_1) * P(Rater_2) + \left(1 - P(Rater_1)\right) * \left(1 - P(Rater_2)\right) \tag{3.8}$$

For the Feasibility inter-rater scores, percent agreement was calculated as defined in Equation 3.7. This metric was chosen because the technical feasibility ratings tend to be very high. As shown in the next chapter, nearly all concepts score an overall 10. Therefore, the weighted kappa metric becomes very skewed by severely penalizing the few disagreements between raters. Calculating the percent agreement was found to be a better reflection of consistency between multiple raters.

Concepts that were generated in past EED experiments were used as training for new raters. From a set of several hundred concepts, subsets of 10 or 20 concepts were randomly selected to be analyzed by four raters: 1 graduate student and 1 professor each from UT and UMD (referred to below as UT1, UT2, UMD1, and UMD2). After a maximum of 3 rounds of this training, all inter-rater kappa scores were high enough to move forward with the analysis.

The training started with UT1 and UT2. The first round of 10 concepts, evaluated after minimal training of UT1, resulted in a weighted kappa of 0.161. This was clearly far from the goal of 0.6 or higher. After discussion and further training, the next round of 10 concepts resulted in a weighted kappa of 0.565. The third round of 20 concepts brought the inter-rater agreement above the acceptable level, to 0.630. At this point, larger concept groups could be analyzed. The weighted kappa from a set of 47 clocks was 0.939 and it was 0.856 from a different group of 81 concepts. These numbers were excellent and the training was expanded to the two raters from UMD. A group of 20 concepts were randomly selected for UMD1 and UMD2 to rate, and then compared to established ratings from UT2. The results from this initial data set are shown in Table 3.6.

Table 3.6: Initial UMD Weighted Kappas

Originality Weighted Kappas			
	UT2	UMD1	UMD2
UT2	x	0.108	0.320
UMD1		x	0.516
UMD2			x

Since these weighted kappa values were below the desired minimum of 0.6, further training was conducted. After reviewing discrepancies and emphasizing use of the exemplar list, the same set of 47 concepts used at UT was evaluated by UMD1 and UMD2. From this set, the weighted kappas between all four raters were above 0.7 and all evaluators felt comfortable with the established rating system. The final inter-rater metrics are summarized in Table 3.7.

Table 3.7: Final Weighted Kappas

Originality Weighted Kappas				
	UT1	UT2	UMD1	UMD2
UT1	x	0.939	0.718	0.784
UT2		x	0.705	0.796
UMD1			x	0.708
UMD2				x

3.5.4 Merging Data

As mentioned, each set of concepts from the 2012-2013 academic year was analyzed by two raters: one from UT and one from UMD. Once the individual ratings of the new concepts were complete, the raters made a merged set of data. To make this merged set, any discrepancies in Originality or Feasibility ratings were discussed and resolved. This step in the process was initially time-consuming but the exemplar list greatly helped reduce the time needed to debate individual ratings. In general, the discrepancies tended to differ by only one rating "level" (for example, 5 vs. 7.5 in Originality or 7 vs. 10 in Feasibility) and could be resolved after a brief review. As the quantity of evaluated concepts increased (from fall to spring), the percent of concepts that needed to be discussed dropped significantly.

As an illustration of typical inter-rater numbers, the Fall 2012 UT data set consisting of 336 concepts will be examined. Two raters (UT1 and UMD1) individually rated all concepts. Prior to merging the individual sets, the weighted kappa for Originality between the two evaluators was 0.808, high enough to be considered "excellent [63]." The percent agreement for Feasibility was also very high at 93%. The weighted kappa from UT1 to the merged set was 0.918 and the weighted kappa from UMD1 to the merged set was 0.906. These numbers imply that the merged data set did not skew towards one rater or the other. Rather, the few disagreements that arose were approximately evenly distributed between the two raters.

After the merged set of ratings was agreed upon, the data could be analyzed for trends. With nearly one thousand concepts in the overall data set, differences could be detected between the various age groups at both universities. The results of the data analysis are presented in Chapter 4.

Chapter 4: Results

4.1 General Results

In total, 975 concepts were evaluated for this study: 564 from students at UT and 411 from students at UMD. Of these designs, 449 (46%) were generated by freshmen and 526 (54%) were generated by seniors. The concepts generated in the Fall 2012 semester were from the beginning of the respective courses, so those terms will be used interchangeably throughout this chapter. The Spring 2013 senior concepts are from the end of the semester at both schools, after the students progressed through the courses named in this study. For the senior groups, this means they were exposed to group concept generation techniques and likely conducted an activity similar to the modified 6-3-5 procedure described in Chapter 3 as part of their group projects. The Spring 2013 freshman data is also from the end of the semester at UT but the UMD spring freshman experiments were run throughout the semester. For the seniors, 278 concepts were from the beginning of the semester and 248 from the end of the semester.

The researchers at UMD also gathered 66 concepts from juniors at the end of the semester, the results of which are not included in this report.

Table 4.1 summarizes the number of students and concepts evaluated for the eight different participant groups.

Table 4.1: Number of Students and Concepts per Group

		Fall		Spring		Total	
		Students	Concepts	Students	Concepts	Students	Concepts
UT	Freshmen	*54*	162	*11*	33	*65*	195
	Seniors	*58*	174	*65*	195	*123*	369
UMD	Freshmen	*17*	52	*67*	202	*84*	254
	Seniors	*35*	104	*18*	53	*53*	157
						325	**975**

Table 4.2 summarizes the average Originality, Feasibility, and Innovation Category scores for the eight student groups, with "% High Orig" representing the percentage of concepts that received a 7.5 or 10 for Originality and "% High IC" representing the percentage of ideas that fulfilled three or more Innovation Categories. For each concept, the overall Originality was calculated as the maximum of the feature-level originality scores. The overall Feasibility score was calculated as the minimum feature-level technical feasibility score for each idea. Across all sets of data, the Feasibility scores tended to be very high, averaging 9.77 out of 10. This was due to the fact that there were very few clock ideas generated that would be technically infeasible to produce. Even concepts that might initially look impractical can still merit a Feasibility rating of 10 if the technology exists elsewhere. Similarly, any design that received a zero for Originality would receive a corresponding 10 for Feasibility since there are no new features present.

Table 4.2: Rating Summary

			Avg. Orig.	Avg. Feas.	Avg. # IC	% High Orig.	% High IC
UT	Freshmen	Fall	1.94	9.87	0.96	8%	15%
		Spring	2.05	9.64	0.70	18%	9%
	Seniors	Fall	3.22	9.86	1.58	14%	27%
		Spring	3.67	9.80	1.70	15%	33%
UMD	Freshmen	Fall	2.84	9.46	1.25	15%	23%
		Spring	2.26	9.75	0.96	10%	14%
	Seniors	Fall	3.13	9.67	1.31	17%	17%
		Spring	4.29	9.66	1.85	17%	32%

4.2 Tests of Statistical Significance

Since the number of concepts in the sample groups ranged from 33 to 202, the researchers used the Mann-Whitney U Test [64] to determine if differences in results were statistically significant from group to group. This test is a non-parametric test to compare two independent groups of sampled data and makes no assumptions of distributions. The test was an improvement when compared to the Student t-test, which is commonly used to find statistical significance between sets of data. The Student t-test assumes normal distributions of data and by looking at the rating distribution histograms in Figures 4.7 through 4.10, it can be seen that this assumption is not valid for all groups in this study. The Wilcoxon signed-rank test [65] was also examined. However, that test is a paired data test and proved inconsistent since some student groups had 3 or 4 times the amount of concepts as other groups. Thus, all p-values mentioned throughout this chapter were calculated using the Mann-Whitney U Test.

Table 4.3 provides the p-values for the Originality ratings of all treatment groups, with statistically significant values ($p<0.05$) highlighted. As was mentioned, the Feasibility averages were all very close to 10. Accordingly, there were never any statistically significant differences between groups in terms of technical feasibility and a table of resulting p-values is not included.

Table 4.3: Originality p-value Summary

					p-values							
					UT				UMD			
					FR		SR		FR		SR	
			Count	Avg Orig	Fall	Spring	Fall	Spring	Fall	Spring	Fall	Spring
UT	FR	Fall	162	1.94	x	0.4325	<0.0001	<0.0001	0.0307	0.1635	0.0024	<0.0001
		Spring	33	2.05		x	0.0107	0.0008	0.0869	0.2389	0.0409	0.0001
	SR	Fall	174	3.22			x	0.0485	0.2061	0.0006	0.3409	0.0047
		Spring	195	3.67				x	0.0287	<0.0001	0.0436	0.0668
UMD	FR	Fall	52	2.84					x	0.1038	0.3409	0.0036
		Spring	202	2.26						x	0.0166	<0.0001
	SR	Fall	104	3.13							x	0.0064
		Spring	53	4.29								x

The significant p-values will be discussed in the following sections.

4.3 Freshmen versus Seniors

Overall, the original hypothesis held true that the senior participants were more creative than their freshman counterparts. The average Originality score across all 975 concepts was 2.887 out of 10. Each freshman group from Table 4.2 averaged an Originality score below this overall average (1.94, 2.05, 2.84, 2.26) while each senior group averaged a higher score (3.22, 3.67, 3.13, 4.29). Figures 4.1 through 4.6 illustrate the average Originality, Feasibility, and Innovation Category ratings at the two universities, separated into Fall and Spring semesters. Again, the Fall data represents “beginning of semester” and Spring represents “end of semester” at both schools, with the exception of the Spring UMD Freshman data being a mixture of beginning and end of semester data. The following sections will discuss the findings from each school in more depth.

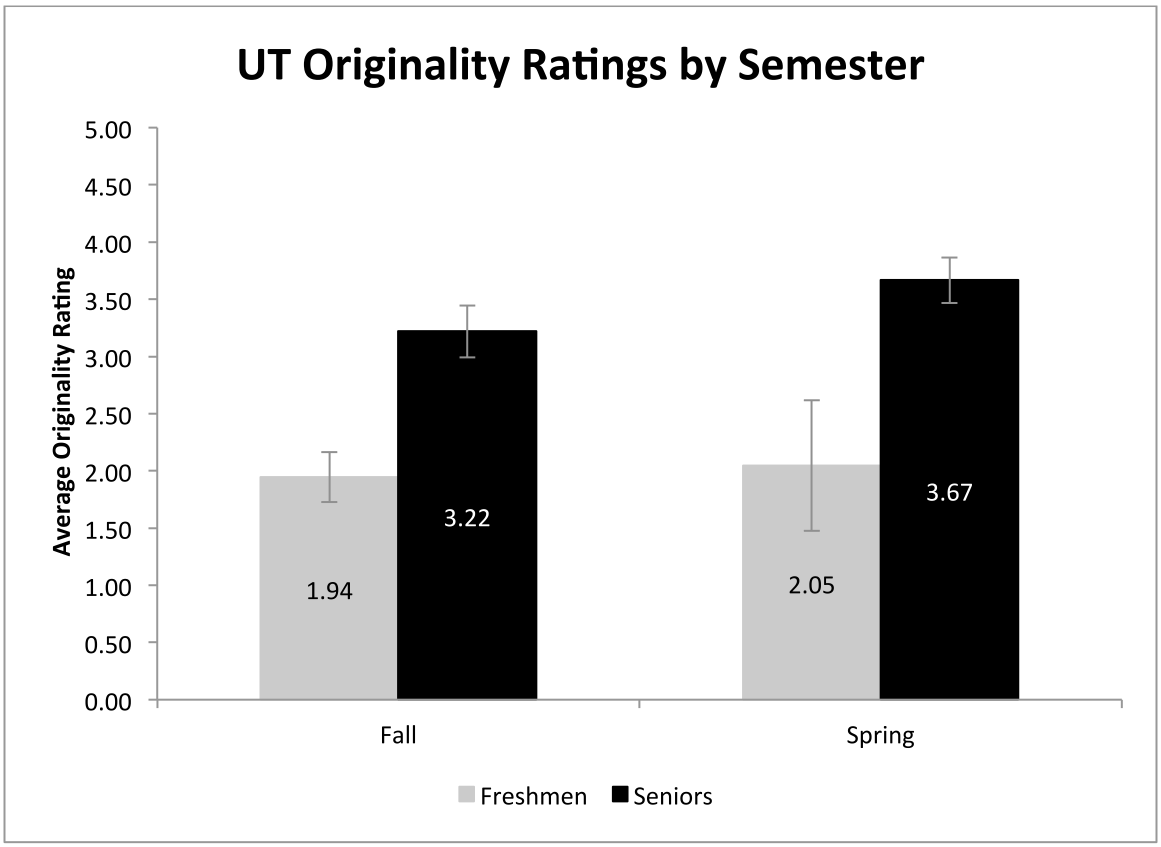

Figure 4.1: UT Originality Ratings

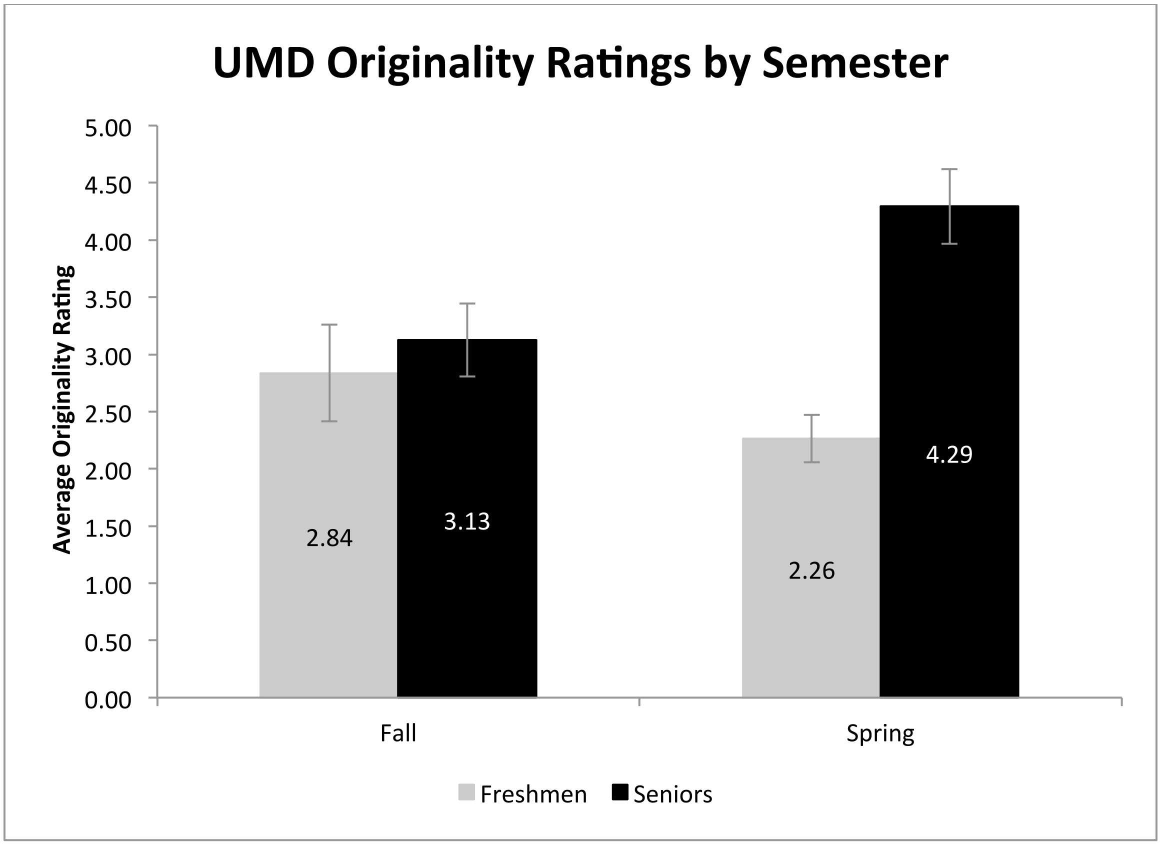

Figure 4.2: UMD Originality Ratings

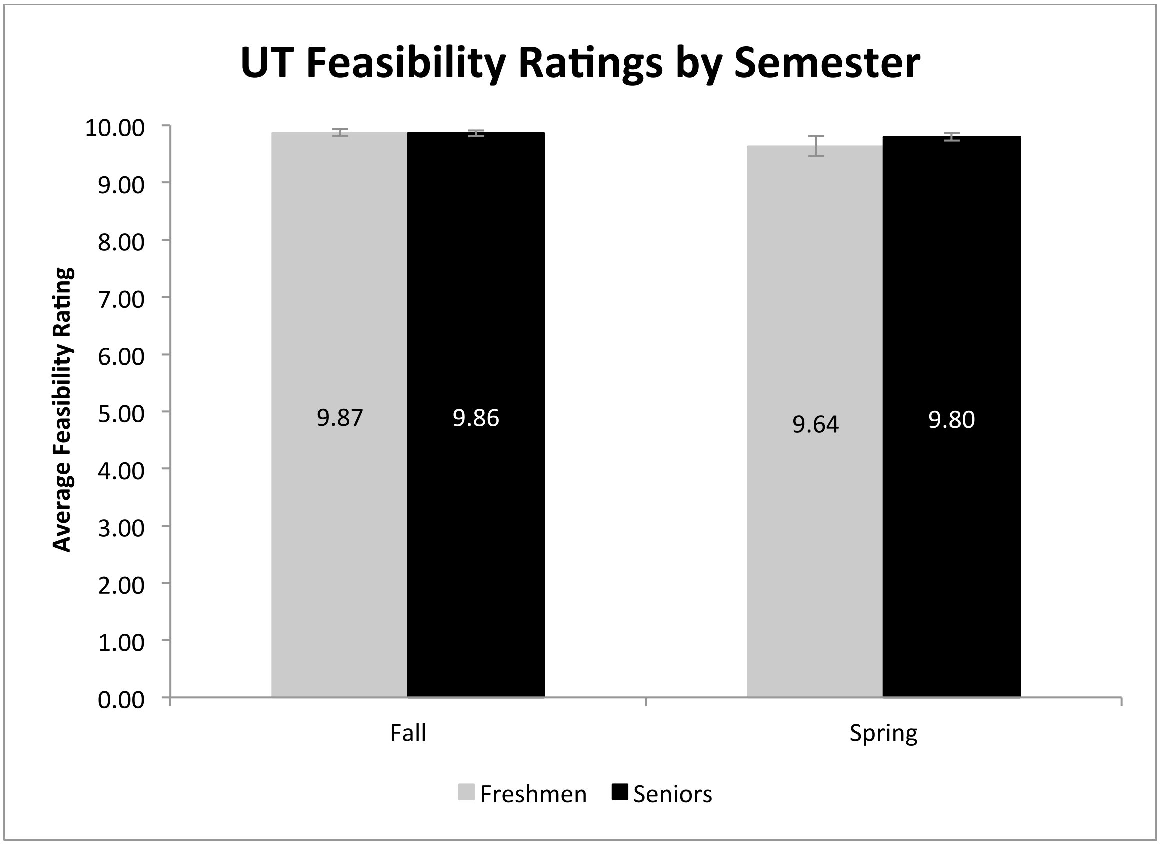

Figure 4.3: UT Feasibility Ratings

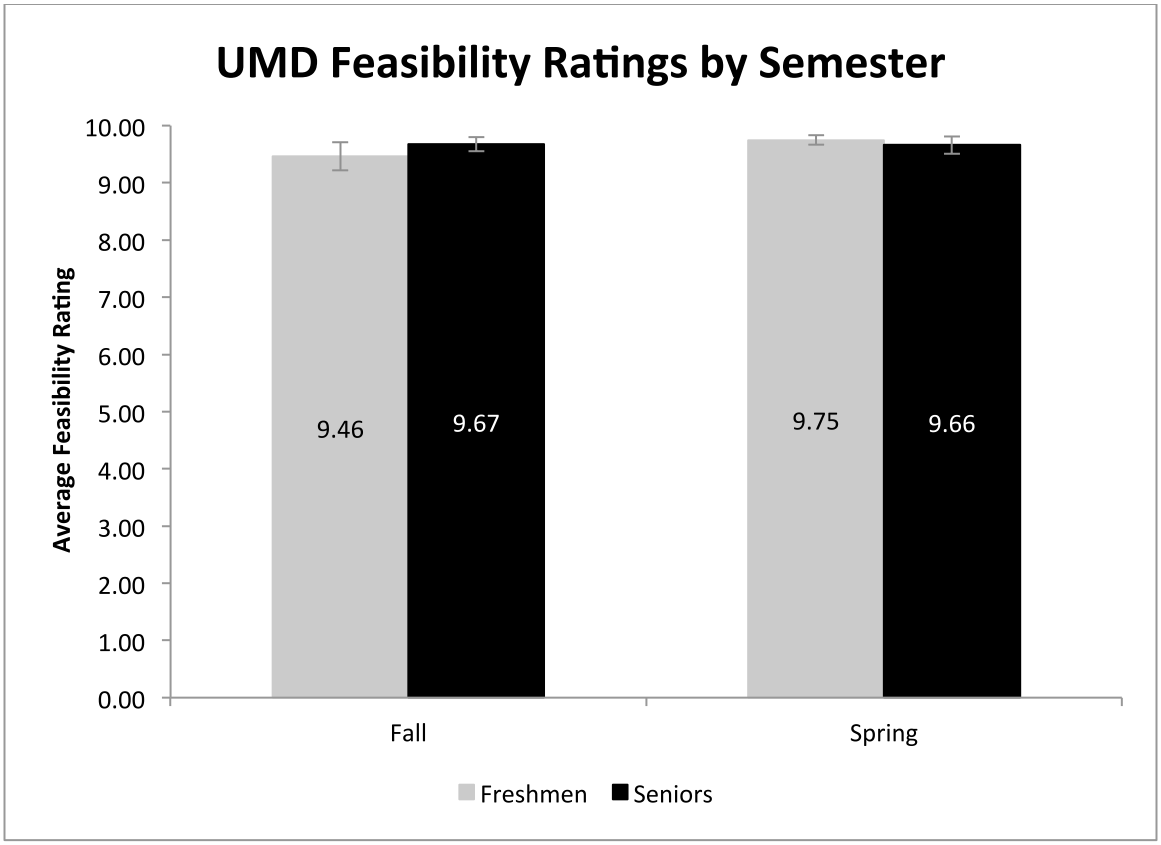

Figure 4.4: UMD Feasibility Ratings

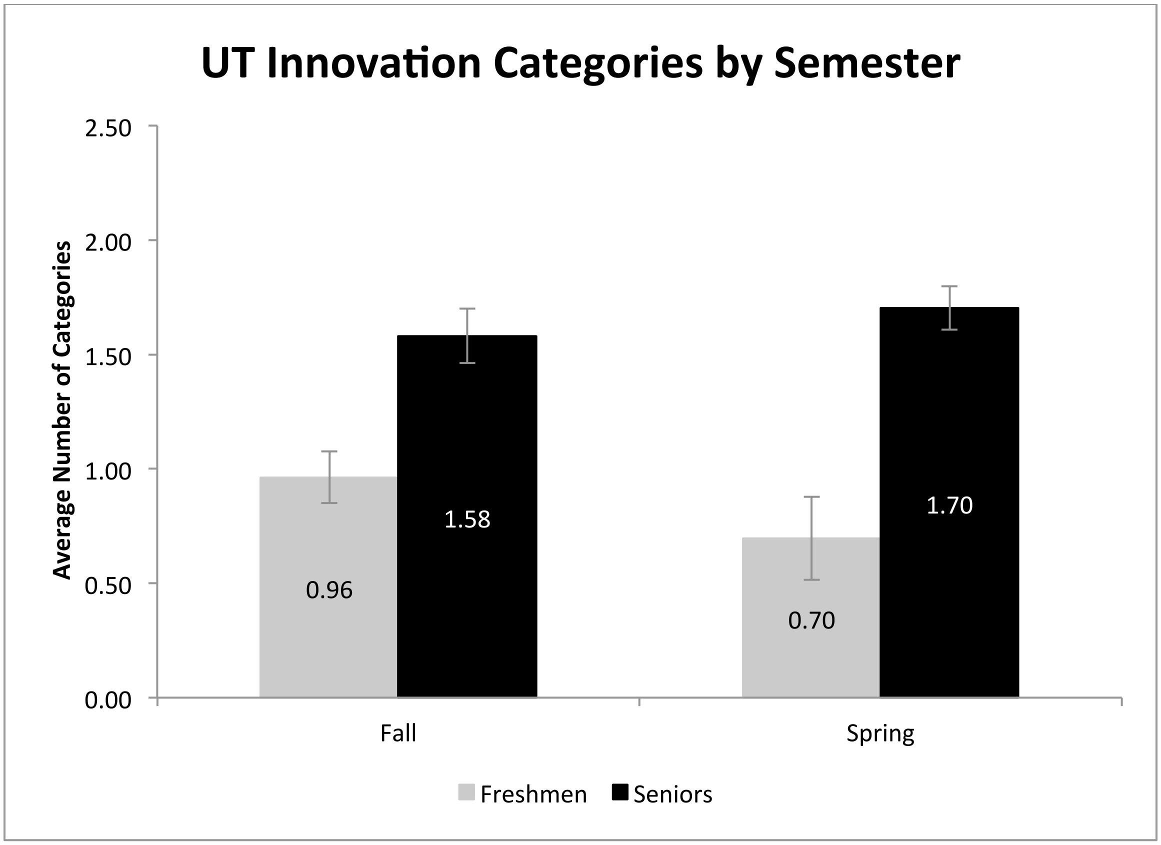

Figure 4.5: UT Innovation Category Ratings

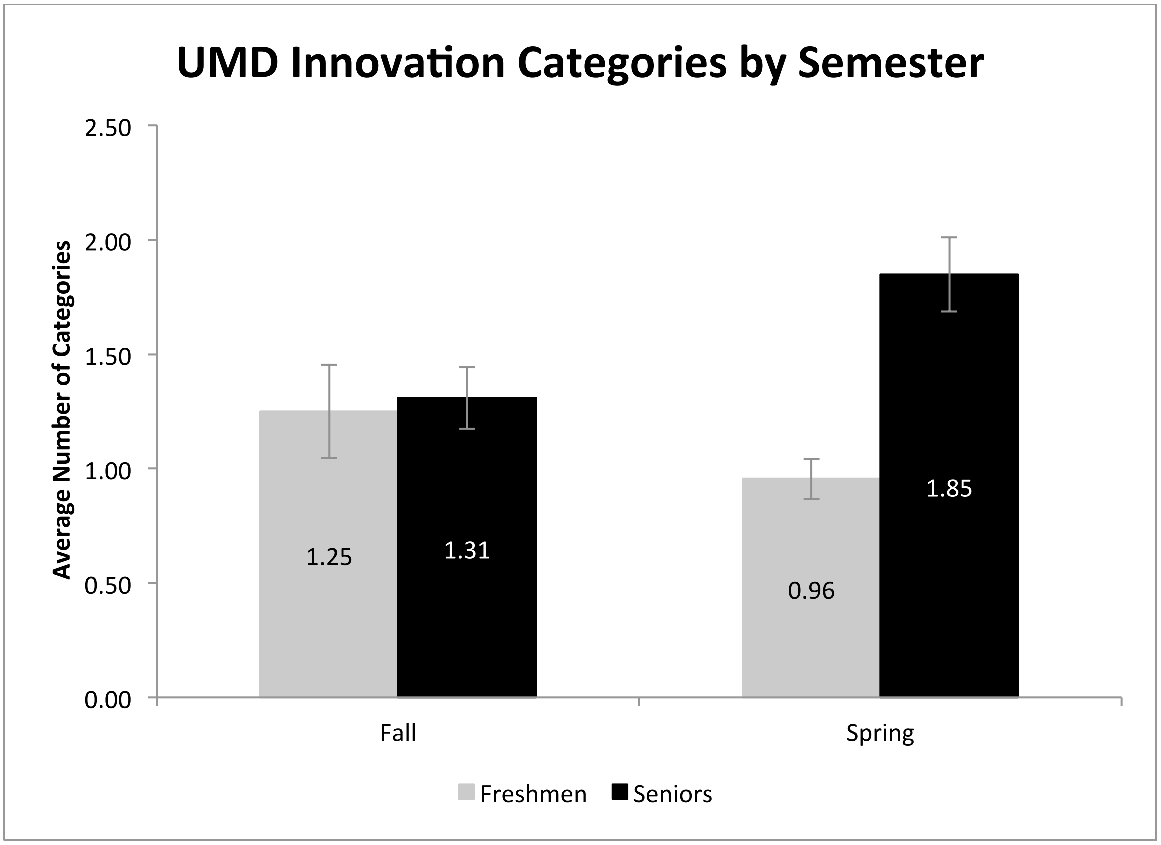

Figure 4.6: UMD Innovation Category Ratings

4.4 UT Results

From the 2012 (Fall) UT data, the freshmen averaged 1.94 for Originality, 9.87 for Feasibility and 0.96 Innovation Categories. The seniors averaged 3.22 for Originality, 9.86 for Feasibility and 1.58 Innovation Categories. In the 2013 (Spring) data, the freshmen averaged 2.05 for Originality, 9.64 for Feasibility and 0.70 Innovation Categories. The seniors averaged 3.67 for Originality, 9.80 for Feasibility and 1.70 Innovation Categories. The Originality scores from the spring senior group (end of semester) were statistically significantly higher than the fall senior group (beginning of semester), with a p-value of 0.049. These results imply that progressing through the Design Methodology class, with exposure to concept generation techniques, increases student innovation capabilities. Since the two groups have statistically significant differences, they were not combined into one large "all-senior" set of data. Conversely, the p-value between the fall and spring freshmen groups was 0.433, showing no significant increase in creativity after completing the first-year course.

Each semester, the seniors significantly outperformed the freshmen in the Originality rating. From the fall data, the p-value between the seniors and freshmen was less than 0.0001. In the spring, the p-value was 0.0008, again showing statistical significance between the groups.

When drawing conclusions from large sets of data, studying only averages can be misleading. However, it can be shown that the seniors' higher averages were not due to a large number of highly rated ideas averaged with a large number of zero-rated ideas. Rather, they performed better across all levels of Originality. From the Fall data distribution shown in Figure 4.7, seniors have a higher percentage of concepts in every non-zero Originality category. From the same set of experiments, more than half of the freshman concepts earned a rating of zero. In the Spring distribution shown in Figure 4.8, the freshmen again accounted for a disproportionate amount of zero scores.

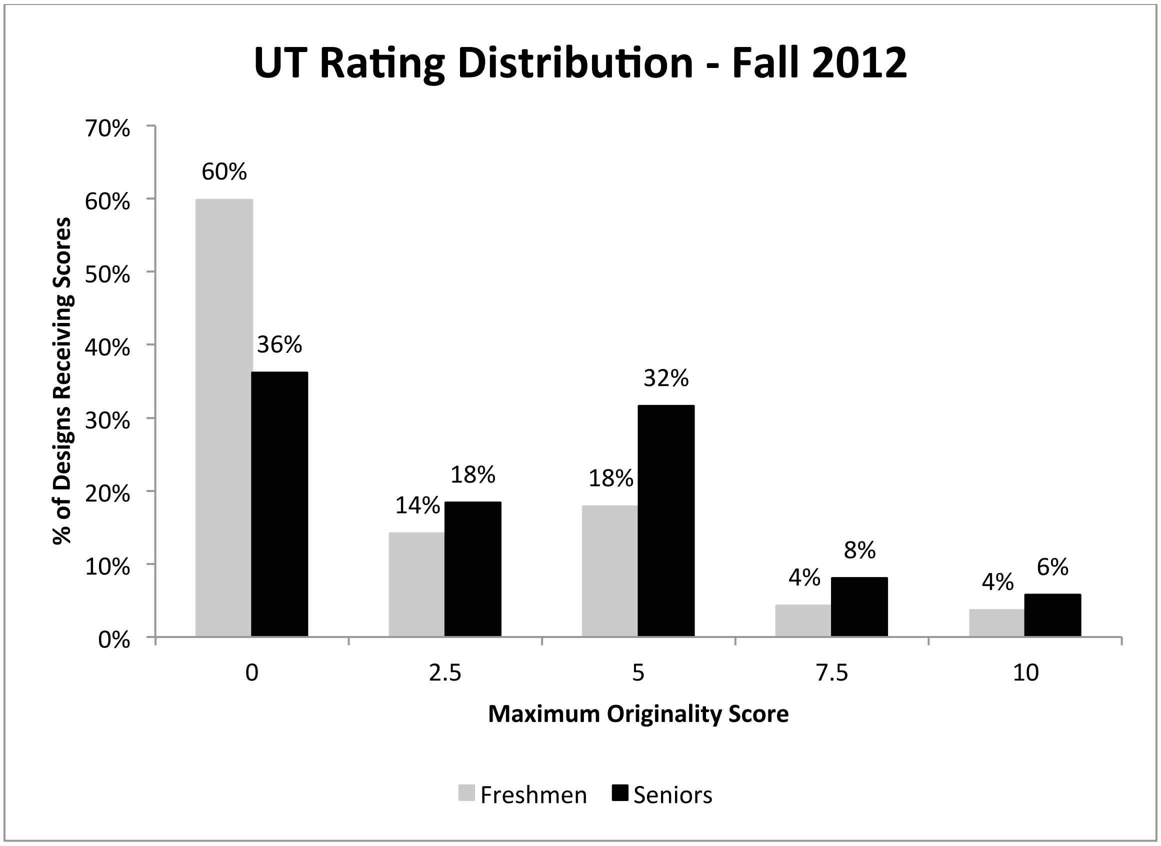

Figure 4.7: UT Fall Originality Rating Distribution

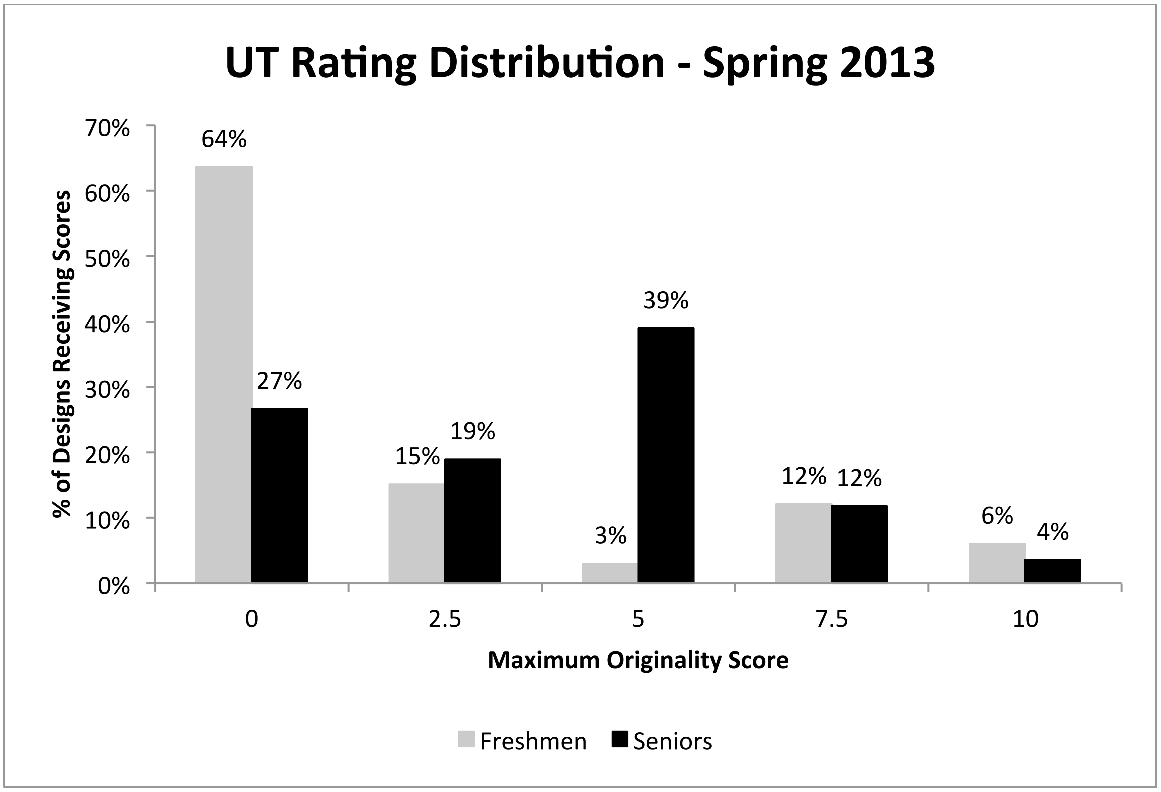

Figure 4.8: UT Spring Originality Rating Distribution

4.5 UMD Results

From the 2012 (Fall) UMD data, the freshmen averaged 2.84 for Originality, 9.46 for Feasibility and 1.25 Innovation Categories. The seniors averaged 3.13 for Originality, 9.67 for Feasibility and 1.31 Innovation Categories. In the 2013 (Spring) data, the freshmen averaged 2.27 for Originality, 9.75 for Feasibility and 0.96 Innovation Categories. The seniors averaged 4.29 for Originality, 9.66 for Feasibility and 1.85 Innovation Categories. Much like at UT, the Originality scores from the spring senior group were statistically significantly higher than the fall senior group, with a p-value of

0.006. Again, the data sets could not be combined and were analyzed by semester. The p-value between the freshmen and seniors for the Fall was 0.34, implying no significant difference between the groups. Figure 4.9 supports this finding, with similar percentages across all Originality ratings for the two groups.

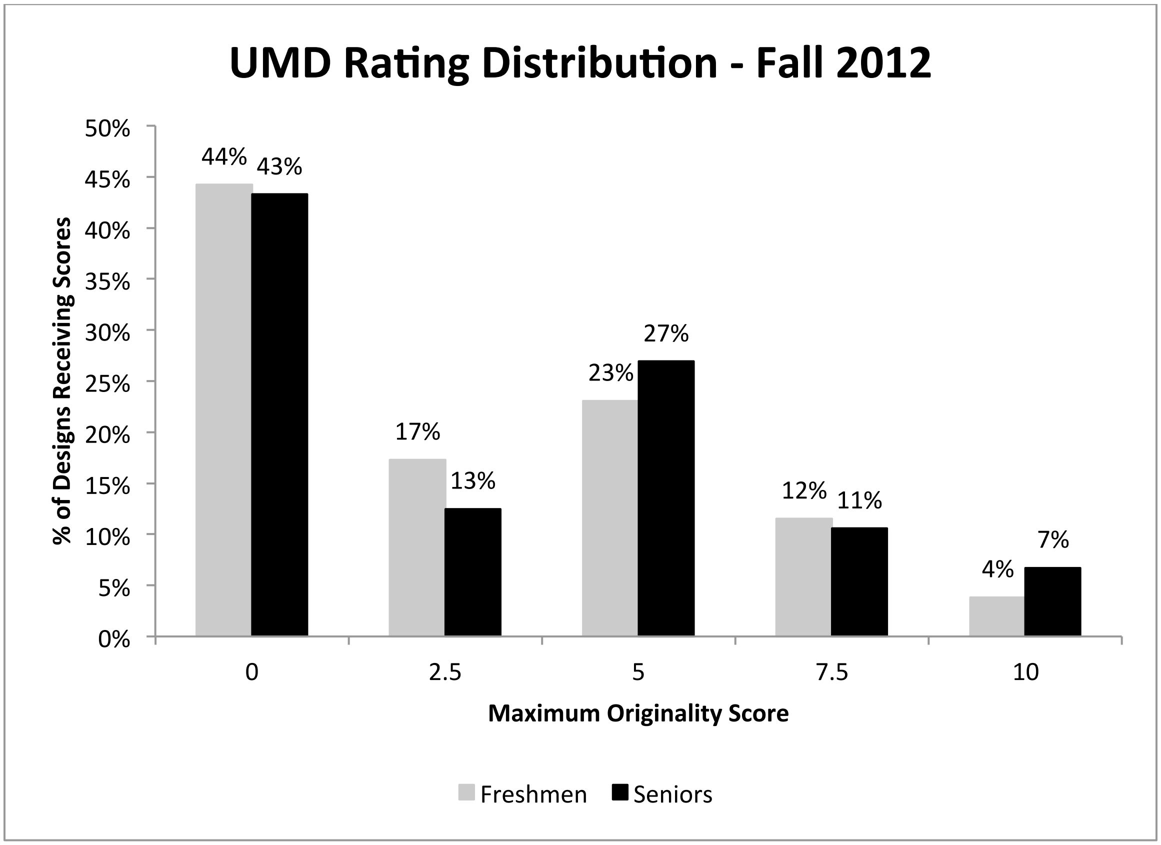

Figure 4.9: UMD Fall Originality Rating Distribution

However, the Spring UMD data, shown in Figure 4.10, displays a much different result. The seniors clearly outperformed the freshmen (p-value <0.0001) with over half of the freshman concepts earning zero points for Originality.

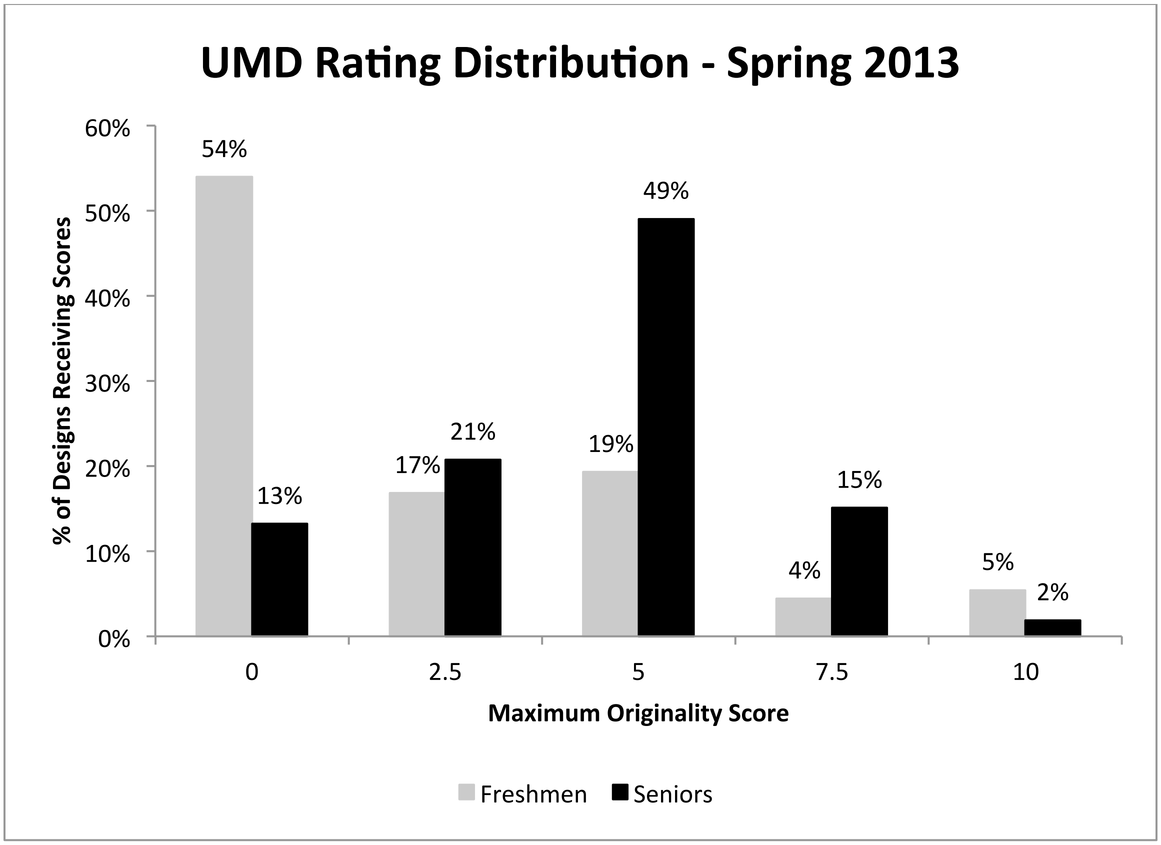

Figure 4.10: UMD Spring Originality Rating Distribution

Interestingly, these results contradict the trend of the results found by Genco *et al*. [19] in experiments conducted at UMD in 2011. This contradiction was possibly due to the refined exemplar list and more consistent system of rating used in the current study. Looking at the older data, some previous raters were much more generous overall in awarding Originality points, especially in the "Additional Uses" category. In the improved rating system, no points were given if the additional function described in the clock concept was not related to waking up the user in some way. Also, due to the prevalence of smartphones, points were no longer awarded for phone-like alarms or apps for phones.

4.6 Surveys

As was described in Chapter 3, the student participants were asked to complete a survey at the end of the experiment. The students were asked to circle their class rank (Freshman, Sophomore, Junior, or Senior) and also their grade point average from a given set of ranges. The survey also included statements for the student to assess his or her abilities, such as "I am good at engineering design" or "I am good at evaluating and testing a design." A blank survey with the full list of statements can be found in Appendix B.

For analysis purposes, the students' GPA and survey responses were converted to numeric values based on scales determined at UMD as shown in Tables 4.4 and 4.5.

Table 4.4: Survey Response Conversion

Value	Response
5	True
4	Somewhat True
3	Not True or False
2	Somewhat Not True
1	Not True

Table 4.5: GPA Conversion

Value	Response
13	3.75-4
12	3.5-3.74
11	3.25-3.49
10	3-3.24
8.5	2.5-2.9
6.5	2.0-2.4
5	<2.0
0	NA or Blank

At UT, the majority of the freshman participants answered "NA" for their GPA since they had just started their academic careers. Therefore, there was not a meaningful set of data to analyze with respect to GPA for the freshmen. However, there was ample data from the seniors to examine.

There was little correlation found between a student's reported GPA and his or her average originality score. In this analysis, the average originality score was calculated from the three concepts that each student initiated. As seen in Figure 4.11, a scatter plot of GPA vs. Average Originality shows a wide dispersion of data, where some students with a high GPA averaged low originality and some of the highest originality scores came from average GPA's.

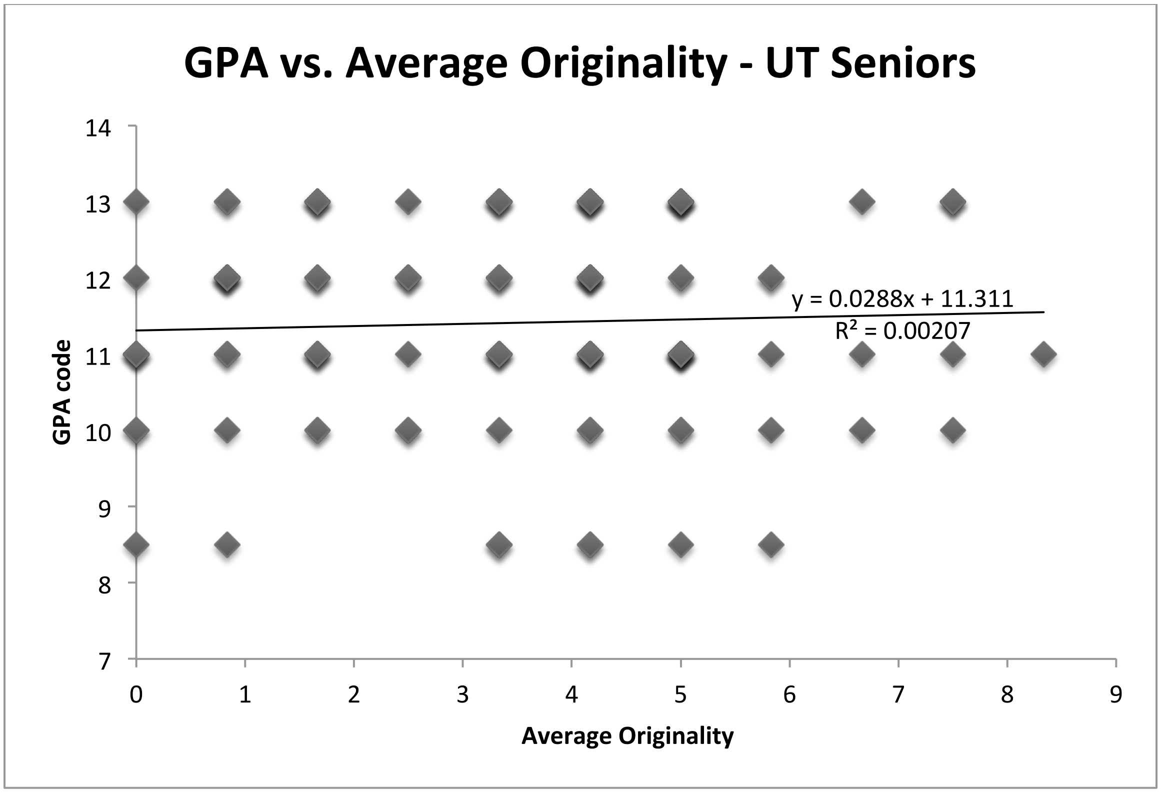

Figure 4.11: GPA versus Average Originality Rating

The survey responses were analyzed for both freshmen and seniors at UT. The researchers divided the students into three levels to compare responses. Those students who had a maximum originality score of 7.5 or 10 were labeled “High Originality,” those with a maximum originality of 2.5 or 5 as “Medium Originality,” and those with a maximum originality of 0 as “Low Originality.” The responses (using the values in Table 4.4) are summarized in Tables 4.6 to 4.9, with each survey statement represented by the corresponding letter code as seen in Appendix B.

Again, it was difficult to draw any significant conclusions from the student responses. Part of the difficulty was that across all students, the average response was 3.96 with a standard deviation of 0.86. Thus, most of the responses were centered on "Somewhat True," with almost zero responses of "Not True." This resulted in minimal variation in averages between groups.

Table 4.6: Average Student Responses

Overall Average												
	A	B	C	D	E	F	G	H	I	J	K	COUNT
Freshman	3.95	4.17	3.95	3.94	4.00	3.75	3.52	3.91	3.85	3.66	4.03	65
Senior	4.02	4.18	4.20	4.15	3.98	3.98	3.55	4.03	3.97	3.83	4.12	123

Table 4.7: Average Student Responses, Low Originality

Average if Max Originality = 0												
	A	B	C	D	E	F	G	H	I	J	K	COUNT
Freshman	3.90	4.20	3.80	3.60	3.75	3.85	3.65	4.00	3.90	3.60	3.55	20
Senior	4.27	4.18	4.00	3.91	4.18	4.36	3.82	4.36	4.27	3.82	4.00	11

Table 4.8: Average Student Responses, Medium Originality

Average if Max Originality = {2.5, 5}												
	A	B	C	D	E	F	G	H	I	J	K	COUNT
Freshman	3.87	4.10	3.97	4.10	4.07	3.53	3.47	3.83	3.73	3.77	4.20	30
Senior	4.03	4.20	4.23	4.15	4.10	3.92	3.44	4.06	4.03	3.87	3.99	71

Table 4.9: Average Student Responses, High Originality

Average if Max Originality = {7.5, 10}												
	A	B	C	D	E	F	G	H	I	J	K	COUNT
Freshman	4.20	4.27	4.13	4.07	4.20	4.07	3.47	3.93	4.00	3.53	4.33	15
Senior	3.93	4.15	4.20	4.20	3.71	3.98	3.68	3.90	3.78	3.76	4.39	41

Looking at the responses of seniors highlights the inconsistencies in the students' self-assessments when compared to the maximum originality scores their concepts earned. Interestingly, the statement in which the High Originality group rated themselves much higher than the Low Originality students was in response to statement K, "I am creative." However, the Low Originality group rated themselves higher than the High Originality group in 8 of the 11 statements, including "I am good at engineering design" and "I am good at communicating my design solution," two statements that directly apply to this design exercise. These self-survey results seem to contradict the scores given by the impartial raters. For the freshmen, the High Originality group was more consistent than the seniors at rating themselves higher than the Low Originality group, but there were still some prompts (H and J) where the reverse was true. Overall, there was little evidence to suggest that the student responses to the self-efficacy survey were accurate predictors of their originality scores in this design exercise.

4.7 Discussion

The researchers further analyzed the UT data sets in an effort to observe qualitative trends between the freshman and senior concepts. In general, the freshmen concepts tended to have more text while the seniors seemed more willing to sketch their ideas. While the source of this discrepancy is unclear, one speculation is that the seniors have had multiple classes with their teammates and therefore are at least acquainted with each other. They are not as afraid of "looking dumb" to their teammates as the freshmen might be. Likewise, the seniors might be more confident in their engineering abilities having completed a few years of coursework or work experience. Another general trend was that more initial sketches seemed to spur more subsequent sketches. Namely, if the first student drew multiple aspects of a design, the following students were more likely to find at least one feature to build on or improve. On the other hand, students seemed to find a small sketch with mostly text difficult to incrementally improve. This was also evident at UMD where some concepts were *entirely* text. This combination of few initial sketches and subsequent comments instead of sketches made it difficult for the evaluators to award originality points to many of the freshmen designs.

Another trend was that the freshmen tended to phrase their comments as questions rather than crossing out or sketching over a teammate's ideas. For example, they might write, "Where is snooze button?" or, "Can it be plugged in?" instead of drawing their idea for the snooze button or power cord. Again, this difference might be a factor of confidence or familiarity with teammates. The freshmen might not want to seem insulting by being overly critical of a teammate's design. Another observation that was seen with less frequency was poor time management by the freshmen. In some cases, a freshman student would have one well-developed idea but then struggle to add two more ideas in the allotted 15 minutes. This could be due to poor listening during the explanation of the experiment or a case of design fixation manifesting as a focus on the first idea that came to the student. The seniors seemed to be better paced in the first round and more able to divide their attention to creating three well-developed ideas.

Overall, there seem to be diminishing returns as the number of rounds increases for both freshmen and seniors. For example, by the time the sixth student receives the concepts, the ideas tend to be very developed or so simple that the student cannot think of anything new to add. At that point they either do nothing or they start to add outlandish features that they find funny since there is no punishment for not contributing anything useful. On one hand, "funny" ideas can lead to creative ideas but the raters decided to ignore notes that were clearly facetious.

At UMD, the senior design class seems to be having a very significant effect on student creativity. At UT, the seniors were statistically significantly more original than the freshmen each semester but at UMD that difference was only significant in the spring, after the students had completed their respective design courses. At both schools, the end-of-semester seniors were statistically higher than the beginning-of-semester seniors. These trends imply that the innovation capabilities of students at UMD, while continually increasing, do not make a significant "jump" until after completing the senior design course.

It is difficult to draw many clear conclusions in a comparison of the two institutions in this study. Since no freshman groups had statistically significant changes between semesters, it is possible to group those concepts by school (into "UT Freshmen" and "UMD Freshmen"). The UT freshmen have a global Originality average of 1.96

while the UMD freshmen average 2.38. This implies that the UMD freshmen start their academic careers at a slightly higher level of creativity. By the time they begin their senior year, their average originality has increased 64% at UT and 31% at UMD. At the end of their senior year, it has increased again by 14% at UT and 37% at UMD, leading to overall increases (from the respective freshmen values) of 87% at UT and 80% at UMD. These numbers imply that the mechanical engineering curricula at both schools are qualitatively meeting the goal of increasing student innovation capabilities. However, since there is no numeric goal for the originality metric, it is difficult to quantify if the courses need to do more. When comparing the end of semester data for the seniors at the two schools, the UMD group has a higher average, but not to a statistically significant degree.

Caution should be taken when analyzing these data sets, as there are a few groups where the sample size is quite small. Specifically, the Spring UT Freshman group (11 students), the Fall UMD Freshman group (17 students) and Spring UMD Senior group (18 students) had fewer participants when compared to the other five groups, which ranged from 35 to 67 students. While the researchers were able to draw some preliminary conclusions from the acquired data, further investigation is recommended to obtain more reliable sample sizes. Small sample size might have been a contributing factor that led to misleading trends in previous research at UMD [19] as well.

4.8 Possible Noise Factors

A study such as this with human participants has many possible sources of error. There are nearly infinite external factors that might influence the results. The researchers tried to control as many parameters of the experiment as possible but the composition of the small groups was randomly determined and can have significant influence on the outcome. For example, if a 6-3-5 group has one or two very creative or "star" students they can significantly raise the group average by contributing original ideas to their 3 original sketches and also to all of their teammates' ideas. This potential issue would be mitigated if all sample sizes were very large, but some of the student groups in this study (such as Spring UT Freshmen) are relatively small. Although the participants were asked at the end of the session not to discuss the content of the experiment with classmates, it is still possible they mentioned to others that the product was an alarm clock or other details of the experiment.

Another possible source of error is the motivation for the students to participate. Both classes from which the students were solicited at UT are required courses in the mechanical engineering curriculum, although this was a voluntary experiment. Since the population of these classes represents many interests or specialties, some of the students simply are not interested in design. However, it can be argued that those students would tend not to volunteer for an activity such as this. The senior-level students were offered individual extra credit in exchange for participation. This could either appeal to high-achieving students eager to accept any chance to raise their grade or to low-achieving students looking for an opportunity to "catch up" to classmates. The freshmen were offered a free dinner in exchange for participating, which might appeal to any type of student.

A factor that can account for higher average senior-level scores is that of freshman students dropping out of the mechanical engineering department. By the time the students reach the senior-level design class, those students who failed previous classes or decided not to pursue engineering have transferred to other departments or left the university. The freshman-level class still includes these students. It would also be interesting to see if there is a difference in the type of student who takes the senior-level class in the spring instead of the fall. With rare exceptions, the class is always taken in the penultimate semester. If those students are taking the class in the spring because they have participated in an internship instead of classes, they might have more developed concept generation skills. On the other hand, they might be taking the class in the spring because they had to repeat a previous class and are now behind in their schedule. Finally, most of the experiments were conducted in the afternoon and evening but a few were run in the morning. It is possible that the time of day had an effect on the students' output, depending on their level of alertness.

Chapter 5: Conclusion

5.1 Summary of Work

Experiments were conducted at the University of Texas at Austin and the University of Massachusetts Dartmouth in an effort to investigate the innovation capabilities of mechanical engineering students. Student volunteers participated in a sketch-based group concept generation exercise to redesign an everyday electromechanical product. Nearly one thousand concepts for next-generation alarm clocks were generated by a combination of freshmen and seniors at both schools. The concepts were then rated for originality, technical feasibility, and innovation characteristics. At both universities, the senior-level participants outperformed their freshman-level counterparts in concept originality ratings without sacrificing technical feasibility. The students also completed a self-efficacy survey at the end of the experiment. There was little correlation found between the self-assessments and the impartial evaluator ratings. Many of the seniors who earned zero originality points rated their abilities higher than those students who scored highly for originality.

5.2 Potential Impacts of the Research

Overall, the results of the study show that mechanical engineering students are more creative at the end of their senior year than they are as freshmen. These results imply that the mechanical engineering curricula are having a positive effect on student creativity and not forcing students into a certain way of thinking. At both schools, the senior group at the end of the semester showed statistically significantly higher originality scores than the senior group selected at the beginning of the semester. This would suggest that the senior-level Engineering Design courses are fulfilling their goal of increasing student creativity at the selected universities to a certain degree. For the freshmen participants, there was no significant change in results when varying the time of semester that the data was acquired.

This research provides a foundation that can be expanded nationally to other universities. An identical procedure could easily be followed elsewhere with minimal training and financial requirements. It would be interesting to see if the same trends that were seen in this study still hold at universities of different sizes and in varying geographic locations. Many variables could be studied, such as public versus private universities, small versus large class sizes, or research-focused versus teaching-focused schools.

As described, the metrics measure the originality, technical feasibility, and innovation characteristics of the clock concepts. These metrics worked well for this study but they are not all encompassing. For example, they do not capture the aspect of marketability. In the study, a vibrating and tilting bed is both original and feasible but without drastic cost-reduction measures, probably will not be commercially viable when competing with present alarm clocks. The metrics also do not attempt to quantify any of the analytical skills that students acquire throughout their academic careers, an important aspect of the mechanical engineering curriculum. Therefore, results of this study should be considered along with many other factors when considering curriculum changes.

5.3 Future Work

This study can be taken in many possible directions for future work. One continuation of the study would be to repeat the same experiment at UT and UMD in three years when the current freshmen are seniors. Although the study is anonymous, the overall makeup of the future group would hopefully have the same demography as the group used in this study. This follow-up analysis would eliminate the possibility of having an exceptionally bright class and would give a clear indication of whether the students' innovation levels have increased after passing through the mechanical engineering curriculum. When performing this follow-up round, it would be helpful to account for the rate of students dropping out of the program, either to transfer to another major or to leave the university entirely.

An aspect of the experiment that would allow the study to easily expand to other universities would be to automate the concept evaluation process. It is very labor-intensive to review every individual concept with multiple raters. As it stands, the process to achieve acceptable agreement between raters is well established and successfully worked in this study. However, in an ideal scenario, the scanned concepts could be run through some sort of feature-recognition software to detect concept qualities and assign them the appropriate rating. This automation could take on other forms, such as having the designs posted online and having the students rate the designs themselves after being shown a few example evaluations. Or the concepts could be put through a service such as Amazon Mechanical Turk, an "artificial artificial intelligence marketplace" where human workers complete tasks that necessitate human intelligence.

In this scenario, some preliminary work would need to be done to ensure the online evaluators' ratings agree with the trained researchers. As an alternative, if a selection of concepts could be analyzed and proven to be representative of the overall group, the necessary evaluation time would decrease appreciably.

One potential improvement to the evaluation process would be to select a different method of rating technical feasibility to output more useful data. The feasibility scores in this study tended to be very high, resulting in little difference between participant groups. However, this may be a factor of the chosen product being an alarm clock. In similar studies using a litter collection system, there was a wider range of feasibility scores.

It may be valuable to create a set of multiple products that can be used in the small group experiments. This would allow the same group of students to be tested at multiple stages of their academic career. Likewise, the same group could be tested at the beginning and end of the same freshman semester and then again as seniors three years later. If the participants are prompted with alarm clocks every time, the results will be greatly skewed. However, if there were a set of established products for the researchers to choose from, the number of data points from a given student group could increase.

Another possible future version of the study is to use different, more open-ended design problems. Although this might make it more difficult to consistently rate and analyze the designs, it might give students a chance to "flex their design muscles" more than when using an alarm clock.

Finally, a noteworthy trend to study would be the movement from traditional, standalone alarm clocks to the use of smartphones or tablets as alarms. One concern with using alarm clocks as the experiment prompt is that other devices are phasing them out of use. Many of the student participants commented that it had been a long time since they had used a clock such as those shown during the introduction of the experiment.

Appendix A: Clock Exemplar List

Mode of Alarm

Originality = 0
- Beeping sound (constant volume or increasing in volume)
- bells ringing, music
- Vibrate
- Nature Sounds
- Low Frequency Flashing lights or display

Originality = 2.5
- Rolling clocks similar to Clocky
- jumping clocks or clocks that move across nightstand by vibrating
- walking legs for movement
- Gradually Lights up (applies to ambient light rather than numbers or other elements of the alarm)
- laser show
- rotating siren horns for pulsing sound
- scent
- confetti

Originality = 5.0
- wake user with high-speed or hot air
- shock to body
- water (e.g., misting, humidifier, water gun)
- vibration (e.g., wristbands, bed)
- puzzle (e.g., press sequence of buttons in order, insert different puzzle shapes to turn off alarm, multiple snooze buttons but only one works)
- helicopter , hovercraft, rocket flight
- punch/slap from alarm
- line up remote with projected image (like Wii; also counts as mode of input)
- projectile (throw against wall to turn off)
- automatically turn the lights on in the room

Originality = 7.5
- flashing projectiles (e.g., lit discs)
- punching bag to be punched with minimum force to turn off alarm
- embed shock in clothing/accessories/socks
- watch that constricts/vibrates/pinches for alarm
- flying helicopter blades
- shock + water + vibrating pillow + fan
- pulls off covers and shakes the bed

Originality = 10
- tilt bed + open blinds + shock
- load cell on floor by bed (must step on it to turn off alarm)
- bed elevated on load cells that sense when user gets out of bed and turn off alarm
- bed deflates
- bed jumps
- refrigerated bed

Display Type

Originality = 0
- Projector (time displayed on wall/ceiling)
- Digital (LED/LCD)
- Analog
- Large numbers (at least 2 inches tall per number)
- Multi-colored (unless to communicate important info such as time remaining before wakeup time)
- Back-lit
- Auditory feedback (Time is read aloud)
- Flashing display
- Glow-in-dark display
- TV screen
- touch screen

Originality = 2.5
- voice-feedback when setting alarm or other functions
- open face, so user can feel time/alarm settings

Originality = 5
- Braille plus open face
- hologram
- turn entire clock face instead of hands moving

Information Shown

Originality = 0
- Current Time
- Time Alarm Set
- Date
- Temperature (both indoor and outdoor)
- Weather
- News, quote, agenda

Originality = 2.5
- synchronize weather with wireless thermometer
- Braille

Originality = 5
- “Okay to Wake" (e.g., alarm clock turns green signifying it is okay for the child to get out of bed)
- weigh you and display your current weight
- adjust wake-up time due to changes in traffic, weather, etc.

Mode of Input

Originality = 0
- Buttons
- Dials (Knobs)
- Atomic Clock (adjusts to time based off radio signals from U.S. atomic clock in Boulder)
- Remote Control
- Touch-screen

Originality = 2.5
- Voice-activated (Basic commands – alarm on-off)
- ergonomic input buttons such as gear-shift style or large rotating wheels
- Braille buttons

Originality = 5.0
- synchronize wakeup to calendar (automatically determine appropriate wake-up time)
- line up remote with projected image to turn off alarm (also listed under mode of alarm)
- password entry to turn off alarm
- shake clock

Originality = 7.5
- punching bag to be punched with minimum force to turn off alarm
- requirement to transport an alarm to a certain location (e.g., back to a stand) to turn it off

Energy Source

Originality = 0
- Batteries
- Electricity
- Wind-up like old analog alarm clocks

Originality = 2.5
- crank-charge (for a battery or capacitor)
- solar powered
- shake to charge

Snooze

Originality = 0
- Button
- Large button

Originality = 2.5
- whole clock snooze
- pull up on handle to snooze

Originality = 5
- only can be pushed when lit up
- unpredictable snooze location (button moves for ex.)
- multiple possible snooze buttons and have to hit the right one

Originality = 7.5
- Rubiks cube to rearrange location of snooze to unpredictable location

Music

Originality = 0
- MP3 player
- iPod
- CD player
- AM/FM Radio
- Cassette Player

Alternative Use

Originality = 0
- Night-light
- USB hookup
- TV
- Digital Photo Album
- coffee maker (timed/alarmed coffee makers have been common for many years)
- cell phone charger
- any alternate use that does not contribute to waking you up

Originality = 2.5
- fan
- attach alarm to toaster or microwave for breakfast foods/smell
- delay/remote car start
- remotely start shower

Shape/Layout

Originality = 0
- Twin Bell Alarm Clock
- Children's Characters (Disney, Hello Kitty, Animals, etc.)
- Collapsible (Travel Type)
- Under the cabinet attachment (for kitchen use)
- wall mount
- key holder
- grandfather clock
- detachable speakers
- folding clock
- any other novelty layout that does not contribute to waking the user (e.g., football shape, UT tower shape)

Originality = 2.5
- whole clock snooze
- creative cord storage (e.g., wind-up)
- unusual layout merged closely with alarm functions (e.g., a car-shaped alarm with flashing lights as headlights and time displayed on windshield)
- attached to bed because it does something to bed (like vibrates it)

Originality = 5.0
- interesting ways to stop rolling clock (e.g., legs jump out to stop rolling)
- set-up like snake light

Originality = 7.5
- flying helicopter blades
- flashing projectiles (e.g., lit discs)
- Rubiks cube to rearrange location of snooze to unpredictable location (also listed under snooze)

Originality = 10
- golf putter or baseball batting requirement (must put golf ball into hole to turn off alarm, for example)
- shock + water + vibrating pillow + fan
- tilt bed + open blinds + shock
- load cell on floor by bed (must step on it to turn off alarm)
- bed elevated on load cells that sense when user gets out of bed and turn off alarm
- earbud in shape of hearing aid that wakes user via wireless signals from base station
- blindfold with vibration incorporated

Appendix B: Innovation Experiment Self-Efficacy Survey

Please complete this survey of your opinions of your innovation abilities and submit it to the investigator before leaving the room. Please identify yourself by number, NOT by name. All research results will be archived by participant number, so that no one can identify you as the participant.

PARTICIPANT NUMBER ______________________

Please circle your current GPA at UT Austin

<2.0 2.0-2.4 2.5-2.9 3-3.24 3.25–3.49 3.5-3.74 3.75-4.0 NA(first semester at UT)

Please circle your official rank at UT Austin

Freshman Sophomore Junior Senior

For each statement below, please pick the answer that best describes your ability

	True	Somewhat true	Not true or false	Somewhat Not true	Not True
A) I am good at engineering design					
B) I am good at identifying a design need					
C) I am good at coming up with new ideas for a design problem					
D) I am good at imagining new ways to solve a problem					
E) I am good at selecting best possible designs					
F) I am good at performing design analysis					
G) I am good at constructing prototypes					
H) I am good at evaluating and testing a design					
I) I am good at communicating my design solution					
J) I am good at redesign					
K) I am creative					

References

[1] T. L. Friedman, "Georgia On My Mind," *The New York Times*, 07-Sep-2008.

[2] B. Obama, "Remarks by the President on the 150th Anniversary of the National Academy of Sciences," 2013. [Online]. Available: http://www.whitehouse.gov/the-press-office/2013/04/29/remarks-president-150th-anniversary-national-academy-sciences. [Accessed: 02-Jul-2013].

[3] B. Obama, "Remarks by the President at The National Academy of Sciences Annual Meeting," 2009. [Online]. Available: http://www.whitehouse.gov/the_press_office/Remarks-by-the-President-at-the-National-Academy-of-Sciences-Annual-Meeting. [Accessed: 02-Jul-2013].

[4] D. F. Heany, "Degrees of Product Innovation," *Journal of Business Strategy*, vol. 4, pp. 3–14, 1983.

[5] A. D. Smith, "Competitive approaches to new product development: A comparison of successful organizations in an unstable economic environment," *Team Performance Management*, vol. 17, no. 3/4, pp. 124–145, 2011.

[6] D. Hoonsopon and G. Ruenrom, "The Impact of Organizational Capabilities on the Development of Radical and Incremental Product Innovation and Product Innovation Performance," *Journal of Managerial Issues*, vol. 24, no. 3, pp. 250–276, 2012.

[7] D. H. Henard and P. A. Dacin, "Reputation for Product Innovation: Its Impact on Consumers," *Journal of Product Innovation Management*, vol. 27, no. 3, pp. 321–335, May 2010.

[8] D. Horn and G. Salvendy, "Measuring Consumer Perception of Product Creativity: Impact on Satisfaction and Purchasability," *Human Factors and Ergonomics in Manufacturing*, vol. 19, no. 3, pp. 223–240, 2009.

[9] R. Cooper, *Product Leadership*. New York: Basic Books, 2005.

[10] A. E. Green, M. S. Cohen, J. U. Kim, and J. R. Gray, "An explicit cue improves creative analogical reasoning," *Intelligence*, vol. 40, no. 6, pp. 598–603, Nov. 2012.

[11] N. Cross, H. Christiaans, and K. Dorst, "Design Expertise Amongst Student Designers," *Journal of Art & Design Education*, vol. 13, no. 1, pp. 39–56, 1994.

[12] C. J. Atman, J. R. Chimka, K. M. Bursic, and H. L. Nachtmann, "A comparison of freshman and senior engineering design processes," *Design Studies*, vol. 20, no. 2, pp. 131–152, Mar. 1999.

[13] M. D. Lammi and J. S. Gero, "Comparing Design Cognition of Undergraduate Engineering Students and High School Pre-Engineering Students," in *2011 Frontiers in Education Conference (FIE)*, 2011, pp. F4F–1–F4F–6.

[14] Engineering Accreditation Commission, "Criteria for Accrediting Engineering Programs." 2012.

[15] C. L. Dym, A. M. Agogino, O. Eris, D. D. Frey, and L. J. Leifer, "Engineering Design Thinking, Teaching, and Learning," *Journal of Engineering Education*, vol. 94, no. 1, pp. 103–120, 2005.

[16] L. Zeng, R. W. Proctor, and G. Salvendy, "Can Traditional Divergent Thinking Tests Be Trusted in Measuring and Predicting Real-World Creativity?," *Creativity Research Journal*, vol. 23, no. 1, pp. 24–37, Feb. 2011.

[17] J. J. Shah, "Identification, Measurement & Development of Design Skills in Engineering Education," 2013, vol. 49, no. 7, p. 644, Paper Number: DS35_557.1

[18] J. J. Shah, R. E. Millsap, J. Woodward, and S. M. Smith, "Applied Tests of Design Skills—Part 1: Divergent Thinking," *Journal of Mechanical Design*, vol. 134, no. 2, pp. 1–10, 2012.

[19] N. Genco, K. Hölttä-Otto, and C. C. Seepersad, "An Experimental Investigation of the Innovation Capabilities of Undergraduate Engineering Students," *Journal of Engineering Education*, vol. 101, no. 1, pp. 60–81, 2012.

[20] S. S. Condoor, S. R. Shankar, H. R. Brock, C. P. Burger, and D. G. Jansson, "A cognitive framework for the design process," in *ASME DETC Design Theory and Methodology Conference*., 1992, pp. 277–281.

[21] D. G. Jansson and S. M. Smith, "Design fixation," *Design Studies*, vol. 12, no. 1, pp. 3–11, Jan. 1991.

[22] A. T. Purcell and J. S. Gero, "Design and Other Types of Fixation or Is Fixation Always Incompatible with Innovation?" 2006.

[23] J. S. Gero, "Fixation and Commitment while Designing and Its Measurement," *Journal of Creative Behavior*, vol. 45, no. 2, pp. 108–115, 2011.

[24] E. G. Chrysikou and R. W. Weisberg, "Following the Wrong Footsteps: Fixation Effects of Pictorial Examples in a Design Problem-Solving Task," *Journal of Experimental Psychology: Learning, Memory, and Cognition*, vol. 31, no. 5, pp. 1134–1148, Sep. 2005.

[25] P. M. Fitts and M. I. Posner, *Human Performance*. Belmont, CA: Brookes Cole, 1967.

[26] J. Y. Lai, E. T. Roan, H. C. Greenberg, and M. C. Yang, "Prompt versus problem: Helping students learn to frame problems and think creatively," in *Third International Conference on Design Computing and Cognition, 2nd Design Creativity Workshop*, 2008.

[27] J. P. Guilford, *The Nature of Human Intelligence*. New York: McGraw-Hill, 1967.

[28] K. Otto and K. Wood, *Product Design: Techniques in Reverse Engineering: Systematic Design, and New Product Development*. New York: Prentice-Hall, 2001.

[29] G. Pahl and W. Beitz, *Engineering Design-A Systematic Approach*. New York: Springer, 1996.

[30] J. J. Shah, "Experimental Investigation of Progressive Idea Generation Techniques in Engineering Design," in *ASME Design Engineering Technical Conference*, 1998, pp. 1–15, Paper Number: DETC98/DTM-5676

[31] A. Wodehouse and W. Ion, "Augmenting the 6-3-5 method with design information," *Research in Engineering Design*, vol. 23, no. 1, pp. 5–15, Apr. 2011.

[32] J. S. Linsey, "Design-by-Analogy and Representation in Innovative Engineering Concept Generation," University of Texas, 2007.

[33] J. S. Linsey, K. L. Wood, and A. B. Markman, "Modality and representation in analogy," *Artificial Intelligence for Engineering Design, Analysis, and Manufacturing*, vol. 22, no. 02, pp. 85–100, Mar. 2008.

[34] B. T. Christensen and C. D. Schunn, "The relationship of analogical distance to analogical function and preinventive structure: The case of engineering design.," *Memory & Cognition*, vol. 35, no. 1, pp. 29–38, Jan. 2007.

[35] I. Chiu and L. Shu, "Biomimetic design through natural language analysis to facilitate cross-domain information retrieval," *Artificial Intelligence for Engineering Design, Analysis, and Manufacturing*, vol. 21, no. 1, pp. 45–59, 2007.

[36] D. Leonard and J. F. Rayport, “Spark Innovation Through Empathic Design,” *Harvard Business Review*, vol. 75, no. 6, pp. 102–113, 1997.

[37] M. N. Saunders, C. C. Seepersad, and K. Hölttä-Otto, “The Characteristics of Innovative, Mechanical Products,” *Journal of Mechanical Design*, vol. 133, no. 2, pp. 1–9, 2011.

[38] A. B. Vangundy, *Techniques of Structured Problem Solving*. New York: Van Nostrand Reinhold, 1988.

[39] R. Van Der Lugt, “Brainsketching and How it Differs from Brainstorming,” *Creativity and Innovation Management*, vol. 11, no. 1, pp. 43–54, Mar. 2002.

[40] T. Buzan and B. Buzan, *The Mind Map Book*. London, UK.: BBC Books, 1993.

[41] G. S. Altshuller, *Creativity as an exact science: The theory of the solution of inventive problems*. Luxembourg: Gordon and Breach Science Publishers Inc., 1984.

[42] P. Hirsch, B. Shwom, C. Yarnoff, J. Anderson, D. Kelso, G. Olson, and J. Colgate, “Engineering Design and Communication: The case for interdisciplinary collaboration,” *International Journal of Engineering Education*, vol. 17, no. 4–5, pp. 342–348, 2001.

[43] J. S. Linsey, E. F. Clauss, T. Kurtoglu, J. T. Murphy, K. L. Wood, and A. B. Markman, “An Experimental Study of Group Idea Generation Techniques: Understanding the Roles of Idea Representation and Viewing Methods,” *Journal of Mechanical Design*, vol. 133, no. 3, pp. 1–15, 2011.

[44] M. Suwa, B. Tversky, J. Gero, and T. Purcell, “Seeing into sketches: Regrouping parts encourages new interpretations,” *Visual and Spatial Reasoning in Design II*, pp. 207–219, 2001.

[45] M. C. Yang, “Observations on concept generation and sketching in engineering design,” *Research in Engineering Design*, vol. 20, no. 1, pp. 1–11, Dec. 2009.

[46] B. Mullen, C. Johnson, and E. Salas, “Productivity Loss in Brainstorming Groups: A Meta-Analytic Integration,” *Basic and Applied Social Psychology*, vol. 12, no. 1, pp. 3–23, 1991.

[47] J. J. Shah, S. M. Smith, and N. Vargas-Hernandez, “Metrics for measuring ideation effectiveness,” *Design Studies*, vol. 24, no. 2, pp. 111–134, Mar. 2003.

[48] B. A. Nelson, J. O. Wilson, D. Rosen, and J. Yen, "Refined metrics for measuring ideation effectiveness," *Design Studies*, vol. 30, no. 6, pp. 737–743, Nov. 2009.

[49] S. K. Oman, I. Y. Tumer, K. Wood, and C. Seepersad, "A comparison of creativity and innovation metrics and sample validation through in-class design projects," *Research in Engineering Design*, vol. 24, no. 1, pp. 65–92, Sep. 2012.

[50] P. Sarkar and A. Chakrabarti, "Assessing design creativity," *Design Studies*, vol. 32, no. 4, pp. 348–383, 2011.

[51] C. Charyton, R. J. Jagacinski, and J. A. Merrill, "CEDA: A Research Instrument for Creative Engineering Design Assessment," *Psychology of Aesthetics, Creativity, and the Arts*, vol. 2, no. 3, pp. 147–154, 2008.

[52] C. Charyton and J. A. Merrill, "Assessing General Creativity and Creative Engineering Design in First Year Engineering Students," *Journal of Engineering Education*, vol. 98, no. April, pp. 145–156, 2009.

[53] R. Srivathsavai, N. Genco, K. Holtta-Otto, and C. C. Seepersad, "Study of existing metrics used in measurement of ideation effectiveness," in *Proceedings of the ASME IDETC Design Theory and Methodology Conference*, 2010, Paper Number: DETC2010–28802.

[54] A. Bandura, *Self-Efficacy: The Exercise of Control*. New York: W.H. Freeman, 1997.

[55] R. Adams, P. Punnakanta, C. J. Atman, and C. D. Lewis, "Comparing Design Team Self-Reports with Actual Performance : Cross-Validating Assessment Instruments," in *American Society for Engineering Education Annual Conference & Exposition*, 2002.

[56] D. G. Johnson, "An Empirical Study of the Effectiveness of Empathic Experience Design," University of Texas, 2012.

[57] M. N. Saunders, "The Effects of Empathic Experience Design Techniques on Product Design Innovation," University of Texas, 2010.

[58] R. Likert, "A Technique for the Measurement of Attitudes," *Archives of Psychology*, vol. 22, no. 140, 1932.

[59] H. Friedman and T. Amoo, "Rating the Rating Scales," *Journal of Marketing Management*, vol. 9, no. 3, pp. 114–123, 1999.

[60] “Clocky.” [Online]. Available: www.nandahome.com/products/clocky. [Accessed: 15-Jul-2013].

[61] J. Cohen, “A Coefficient of Agreement for Nominal Scales,” *Educational and Psychological Measurement*, vol. 20, no. 1, pp. 37–46, Apr. 1960.

[62] J. Fleiss, *Statistical Methods for Rates and Proportions*. New York: John Wiley and Sons, 1981.

[63] J. R. Landis and G. G. Koch, “The Measurement of Observer Agreement for Categorical Data,” *Biometrics*, vol. 33, no. 1, pp. 159–174, 1977.

[64] H. B. Mann and D. R. Whitney, “On a Test of Whether One of Two Random Variables is Stochastically Larger than the Other,” *The Annals of Mathematical Statistics*, vol. 18, no. 1, pp. 50–60, 1947.

[65] F. Wilcoxon, “Individual Comparisons by Ranking Methods,” *Biometrics Bulletin*, vol. 1, no. 6, pp. 80–83, 1945.

Vita

Paul T. Williams was born in Milwaukee, Wisconsin. He attended high school in Brookfield, Wisconsin, graduating in 2002. He attended Northwestern University in Evanston, Illinois, and graduated in 2006 *magna cum laude* with a Bachelor of Science in Mechanical Engineering. Paul then went to work for HUSCO International in Waukesha, Wisconsin as a Design Engineer. He began his graduate studies at The University of Texas at Austin in August 2011.

Permanent email: paul.t.williams1@gmail.com

This thesis was typed by the author.