SUPPORTING EXAMPLE-BASED IDEATION AND ASSESSMENT PRACTICES IN ENGINEERING DESIGN

BY

SCARLETT R. HERRING

DISSERTATION

Submitted in partial fulfillment of the requirements for the degree of Doctor of Philosophy in Industrial Engineering in the Graduate College of the University of Illinois at Urbana-Champaign, 2011

Urbana, Illinois

Doctoral Committee:

Associate Professor Ali Abbas, Chair
Professor Alex Kirlik, Director of Research
Associate Professor Brian Bailey
Professor Deborah Thurston
Abstract

Over the past decade, examples have become the cornerstone of the design process. In essence, designers are no longer developing ideas from scratch, but instead designing through the synthesis of pre-existing design ideas. The repercussion of this new design paradigm is the quality of the design output has become increasingly dependent on the types of examples designers retrieve for inspiration. Although examples use is pervasive in many design disciplines, little research has been conducted on how and why designers use examples during the design process. This is problematic for the understanding of design activities, as well as for the development of more effective design tools. In addition, example use has not been studied outside of the idea-generation process and thus little is known about what other uses examples could have. For example, one’s ability to identify high quality design examples during the design process has been linked to the designer’s expertise, yet no study to date has explored how examples could be used as an assessment technique.

Therefore, the purpose of this dissertation was two-fold. First, we sought to provide an understanding of how and why examples are used in design practice and how we can better support these methods through the enhancement of existing design tools or the creation of new ones. Next, we sought to understand how ratings of example quality can be used to predict one’s design expertise in evaluating ideas, creating ideas and critiquing ideas. In total, six experiments were conducted. The results revealed several challenges to supporting example usage in design. First, the types of examples collected for inspiration can have a negative impact on the designer’s ability to develop innovative solutions. As such, new computational tools are needed that help individuals collect a diverse example set and develop new design directions. In addition, assessments of student design competence are based primarily on subjective evaluations of student performance by the course instructor, a design expert. However, these experts are subject to cognitive biases based on their own beliefs and expectations. Our example-based assessment method utilizes the Bayesian Truth Serum algorithm which removes instructor bias by taking them out of the evaluation criterion. Our results show a positive correlation with student’s ability to analyze and evaluate design ideas, but not their ability to generate innovative solutions. Implications and future research directions are discussed.
To my Husband
Acknowledgements

I would like to thank my advisors Professors Alex Kirlik and Brian Bailey for the guidance and support they have given me throughout my time at Illinois. I have learned so much from both of them from how to conduct top-tier research to how to navigate the academic job market, to how to become the type of advisor every student wants (one who only expects the best work from their students, pushes them until they get it, but maintains a positive mentor relationship with them the whole time; which I must say is truly a feat). I only hope that you realize how much I truly appreciate the support you have given me.

To my ORCHID lab mates Moushumi, Roshanak, Anbang, Braj, Chia-Chen, Jake and Shamsi who were at one time or another students with me, I want to thank you for helping me with my research and for lending a welcome ear when I needed to chat (and vent!). My time at Illinois would not have been the same without you!

To my mom, dad and sister; thank you for supporting my career path as a professional student and for believing that I would one day actually graduate and get a ‘real’ job. I thank you for your love and support and hey, I finally did it.

Finally, to my husband; Thank you for supporting me throughout my graduate career, for picking me up at Siebel Center at 3 am, for making me coffee to keep me awake and for listening to me go on and on about my research projects. Thank you for helping ‘celebrate’ my successes (and sometimes failures). I am so lucky to have you in my life.
Table of Contents

CHAPTER 1: INTRODUCTION ........................................................................................................... 1
1.1 RESEARCH OVERVIEW AND SIGNIFICANCE ........................................................................... 3

PART I: EXAMPLE-BASED IDEATION PRACTICES ........................................................................ 6

CHAPTER 2: LITERATURE REVIEW: THE DESIGN PROCESS ......................................................... 7
2.1 ENGINEERING DESIGN AND CREATIVE PROCESS ................................................................. 8
2.2 DESIGN PROCESS-ANALYSIS ............................................................................................... 14
2.3 DESIGN PROCESS- IDEA GENERATION ............................................................................... 14
2.4 DESIGN PROCESS- EVALUATION ......................................................................................... 23
2.5 COMPUTATIONAL TOOLS FOR SUPPORTING IDEA GENERATION IN DESIGN ............... 24
2.6 SUMMARY OF ENGINEERING DESIGN LITERATURE .......................................................... 26

CHAPTER 3: EXPERIMENT 1 IDEA GENERATION TECHNIQUES .................................................. 28
3.1 METHODOLOGY ................................................................................................................... 30
3.2 RESULTS ................................................................................................................................ 32
3.3 SUMMARY AND DISCUSSION ............................................................................................... 43
3.4 FUTURE DIRECTIONS ............................................................................................................ 48

CHAPTER 4: EXPERIMENT 2 EXAMPLE USAGE IN DESIGN ....................................................... 50
4.1 METHODOLOGY ................................................................................................................... 51
4.2 RESULTS ................................................................................................................................ 54
4.3 SUMMARY AND DISCUSSION ............................................................................................... 72
4.4 FUTURE DIRECTIONS ............................................................................................................ 76

CHAPTER 5: EXPERIMENT 3 EXAMPLE RETRIEVAL .................................................................... 77
5.1 METHODOLOGY ................................................................................................................... 78
5.2 RESULTS ................................................................................................................................ 86
5.3 SUMMARY AND DISCUSSION ............................................................................................... 99
5.4 SUMMARY PART I ................................................................................................................ 103

PART II: DESIGN SKILL ASSESSMENT .................................................................................... 107

CHAPTER 6: LITERATURE REVIEW DESIGN SKILL ASSESSMENT ............................................... 108
6.1 ENGINEERING DESIGN COURSE STRUCTURE AND LEARNING OBJECTIVES ............. 109
6.2 ASSESSMENT IN ENGINEERING DESIGN ......................................................................... 113
6.3 SUBJECTIVE ASSESSMENT MEASURES IN ENGINEERING DESIGN ............................. 115
# Table of Contents Cont'd

<table>
<thead>
<tr>
<th>Chapter</th>
<th>Title</th>
<th>Pages</th>
</tr>
</thead>
<tbody>
<tr>
<td>7</td>
<td>THE BAYESIAN TRUTH SERUM ASSESSMENT</td>
<td>118-123</td>
</tr>
<tr>
<td>7.1</td>
<td>Bayesian Truth Serum</td>
<td>119</td>
</tr>
<tr>
<td>7.2</td>
<td>BTS Method</td>
<td>120</td>
</tr>
<tr>
<td>7.3</td>
<td>OVERVIEW OF EXPERIMENTS</td>
<td>122</td>
</tr>
<tr>
<td>7.4</td>
<td>SUMMARY</td>
<td>123</td>
</tr>
<tr>
<td>8</td>
<td>EXPERIMENT 4 BTS SKILL ASSESSMENT</td>
<td>124-138</td>
</tr>
<tr>
<td>8.1</td>
<td>METHODOLOGY</td>
<td>124</td>
</tr>
<tr>
<td>8.2</td>
<td>PROCEDURE OVERVIEW</td>
<td>127</td>
</tr>
<tr>
<td>8.3</td>
<td>RESULTS</td>
<td>129</td>
</tr>
<tr>
<td>8.4</td>
<td>SUMMARY AND DISCUSSION</td>
<td>136</td>
</tr>
<tr>
<td>8.5</td>
<td>FUTURE DIRECTIONS</td>
<td>138</td>
</tr>
<tr>
<td>9</td>
<td>BTS AND IDEA GENERATION</td>
<td>139-153</td>
</tr>
<tr>
<td>9.1</td>
<td>METHODOLOGY</td>
<td>140</td>
</tr>
<tr>
<td>9.2</td>
<td>PROCEDURE OVERVIEW</td>
<td>140</td>
</tr>
<tr>
<td>9.3</td>
<td>RESULTS</td>
<td>146</td>
</tr>
<tr>
<td>9.4</td>
<td>SUMMARY AND DISCUSSION</td>
<td>151</td>
</tr>
<tr>
<td>9.5</td>
<td>FUTURE DIRECTIONS</td>
<td>153</td>
</tr>
<tr>
<td>10</td>
<td>EXPERIMENT 6 BTS AND THE DESIGN CRITIQUE</td>
<td>154-161</td>
</tr>
<tr>
<td>10.1</td>
<td>METHODOLOGY</td>
<td>154</td>
</tr>
<tr>
<td>10.2</td>
<td>PROCEDURE OVERVIEW</td>
<td>155</td>
</tr>
<tr>
<td>10.3</td>
<td>EXPERIMENTAL DESIGN</td>
<td>157</td>
</tr>
<tr>
<td>10.4</td>
<td>RESULTS: DOES BTS CORRELATE WITH CRITIQUE QUALITY</td>
<td>158</td>
</tr>
<tr>
<td>10.5</td>
<td>SUMMARY AND DISCUSSION</td>
<td>160</td>
</tr>
<tr>
<td>10.6</td>
<td>SUMMARY PART II</td>
<td>161</td>
</tr>
<tr>
<td>11</td>
<td>CONCLUSION</td>
<td>162-174</td>
</tr>
<tr>
<td>11.1</td>
<td>PART I: EXAMPLE-BASED IDEATION PRACTICES</td>
<td>163</td>
</tr>
<tr>
<td>11.2</td>
<td>SUMMARY PART II: EXAMPLE-BASED ASSESSMENT PRACTICES</td>
<td>166</td>
</tr>
<tr>
<td>11.3</td>
<td>IMPLICATIONS</td>
<td>170</td>
</tr>
<tr>
<td>11.4</td>
<td>LIMITATIONS AND FUTURE DIRECTIONS</td>
<td>174</td>
</tr>
<tr>
<td>12</td>
<td>REFERENCES</td>
<td>175</td>
</tr>
<tr>
<td>APPENDIX A</td>
<td>BAYESIAN TRUTH SERUM SURVEY</td>
<td>184</td>
</tr>
<tr>
<td>APPENDIX B</td>
<td>TRUE/FALSE QUIZ EXPERIMENT 4</td>
<td>185</td>
</tr>
</tbody>
</table>
Chapter 1

Introduction

The field of design has changed from a design from scratch to a design through synthesis environment where designers transform, combine or adapt elements of existing designs in order to generate new ideas (C. Eckert, Stacey, & Earl, 2005). In other words, designers in a variety of disciplines including engineering, architecture and creative design (Web, graphic and industrial) decompose existing design examples in order to identify and reuse interesting components in their own design ideas (Nickerson, 1999; Ward, 1995; Weisburg, 1995). Because design has become so example-centric, the creativity and originality of designers’ outputs are now largely dependent on the quality of examples they retrieve during the design process. In other words, if designers’ collects high quality, innovative design examples they are likely to develop high quality ideas. On the contrary, if designers collect low-quality design examples, their design outputs are likely to be uninspired (A. R. Purcell & Gero, 1996).

Although we know that the types of examples collected during the design process affects the creativity of design outputs, the process has not been widely studied and thus little is known about how designers collect examples during the design process and what types of examples designers choose for inspiration (Tseng, Moss, Caga, & Kotovsky, 2008). Therefore, we are unable to conclude if examples are currently aiding in the design process by promoting cognitive
stimulation and the flow of useful ideas, or if they are actually impeding this process. This is problematic for both the understanding of design activity as well as for developing more effective design tools.

In addition, because of the deficit of knowledge about example-centric design methods, we know very little about what other uses examples may have in professional and educational design practices. For instance, while research has identified a relationship between a designer’s ability to recognize high quality examples and their design knowledge, i.e. the more knowledgeable or experienced the designer the more able they are to identify high quality design examples, no study to date has explored examples as a design knowledge assessment tool (N. Bonnardel & Marmèche, 2004). In other words, this research finding which links one’s design knowledge with their ability to rate example quality could allow us to predict ones’ competence in design through their ratings of example quality. This would extend example usage from purely an inspirational design tool to a design knowledge assessment tool. This new use of examples could be beneficial both for professional design practice as well as design education.

In particular, engineering design education would greatly benefit from the development of an objective assessment of design skill as the role and importance of design in engineering education has changed significantly over the past five decades and instructors are struggling to develop effective assessment methods. Historically, engineering curricula were based on an ‘engineering science’ model in which engineering design is taught only after a solid foundation in science and math (Grinter, 1956). Nowadays, however, design is considered to be the central or distinguishing activity of engineering and thus is being widely integrated throughout engineering education (Kilgore, Atman, Yasuhara, Barker, & Morozov, 2007; Morozov, Kilgore, & Atman, 2007; Simon, 1996). Because of this change, instructors now use a variety of methods to teach students design skills such as design labs, open-ended course projects, and objective
(multiple-choice) tests. While there are a variety of assessments that have been developed, most objective assessments are untested for reliability of validity and thus instructors often have little confidence in their assessment of student competence in design (Davis, Beyerlein, Trevisan, Thompson, & Harrison, 2006). In addition, while projects and labs allow for the development of design skills, subjective evaluations are utilized to evaluate student performance reducing their effectiveness. Therefore, new assessments are needed to provide an objective evaluation of these design skills. Example-centric methods may be able to fill this void in the assessment process.

In sum, while we know example usage is pervasive in a variety of design disciplines, we know very little about how designers retrieve examples and the types of examples they collect for inspiration. This makes it difficult to develop new design methods and design tools that support example-centric design methods. In addition, example usage has only been studied in the context of idea generation and as such no one has identified new uses for examples outside of this setting. Thus, it is impossible to tell how example-centric methods could be used in other design applications, such as the assessment of one’s competence in design.

1.1 Research Overview and Significance

The purpose this dissertation is two-fold. First, we seek to develop an in-depth understanding of example-centric ideation practices in order to develop new design methods and implications for future design tools that better support these methods. Next, we seek to understand how examples can be used as an assessment tool in engineering design education by testing a novel application of the Bayesian Truth Serum, a method for eliciting subjective information when objective truth is intrinsically or practically unknowable (i.e. if an example is useful for a design task). Thus this dissertation is organized in two parts: Part I: Example-Based Ideation Practices
and Part II: Example-Based Assessment Practices. Three experiments were conducted in each part of this dissertation for a total of six studies.

In Part I of this dissertation, three studies were conducted to develop an understanding of existing ideation practices in professional design. These studies focus on both engineering and creative design practices since examples are used in a variety of design disciplines and the generalizability of our results hinges on our understanding of example usage from the entire design community. These studies are based primarily on interviewing and surveying some of the best designers in the country in order to develop a better understanding of the ‘best’ design practices. The implications from Part I of this dissertation seek to impact the development of new example-centric design methods and new design tools that support example-centric ideation methods.

Part II of this dissertation focuses on developing and validating a new objective measure of student design skill for use in engineering design education. In order to accomplish this, three studies were conducted to understand how ratings of example quality, conceptual ideas, and design critiques can be used to identify one’s design skill level. This was accomplished by using a novel application of the Bayesian Truth Serum (BTS), and was studied in a Human Factors Engineering Design course at the University of Illinois. The implications from Part II of this dissertation seek to impact the development of objective measures of student design skill with implications for professional design practice.

The results from Part I and II of this dissertation are intended to make contributions to several key areas of the literature. First, by understanding existing example-based design methods we can develop guidelines for next-generation design tools. This work is also design to develop an understanding of how designers collect examples and the types of examples they collect in order
to identify deficits in these processes and develop new design research directions. Next, by studying how objective assessment techniques can be used to test different types of design skills in engineering design education, we are contributing new knowledge on how objective assessments can be utilized to test more than domain specific knowledge. In addition, by testing a novel application of the Bayesian Truth Serum we are adding new knowledge on the applicability and generalizability of the method. Finally, by studying how these objective measures can be utilized with minimal expert (instructor) input, we are developing new ways assessments can be conducted in engineering design education.

This dissertation is organized as follows: Chapter 2 provides a review of the design process literature base that serves as the foundation for Part I of this dissertation. Chapter 3 through 5 outlines studies completed on supporting example-based ideation practices in design. Chapter 6 and 7 outline literature related to engineering design assessment for Part II of this dissertation. Chapter 8-10 outline three studies completed for Part II on developing new objective assessments of student design skill. Finally, Chapter 11 provides a summary of the dissertation studies and highlights the contributions of this work to the research community.
Part I: Example-Based Ideation Practices
Chapter 2

Literature Review: The Design Process

There are many definitions of what engineering is and what engineers do, but the contemporary view of the engineering profession is that engineers, “scope, generate, evaluate and realize ideas” (Sheppard, 2003). This characterization by Sheppard focuses on how engineers think and embrace the heart of the design process in terms of creation (scoping and generating), assessment, selection (evaluation) and the bringing to life of new ideas (realization) (Dym, 2006). The changing view of the engineering profession to be more design centric brings to light several interesting questions. First, what constitutes the engineering design process and how does this process compare with the creative design process. Next, how does one develop, store and retrieve knowledge to help in the design process. Finally, what difficulties do designers have in the design process, and what types of computational support currently exists to help designers in these processes? The objective of Part I of this dissertation is to provide a better understanding of the design process in order to inform the design of new tools to support example-centric ideation methods. This chapter serves as an outline of previous research aimed at answering the questions outlined above. Following this discussion I will detail three studies completed to identify deficiencies in current design practices and highlight opportunities for new design tools that better support these processes.
2.1 Engineering Design and Creative Process

Creativity is an integral part of the engineering design process as without creativity there is no potential for innovation (T. Amabile, 1996; Mumford & Gustafson, 1988). Therefore, it should come as no surprise that the engineering design and creative design process have many similarities (Howard, Culley, & Dekoninck, 2007). Before we can understand the similarities between these processes and how we can better support them, we first must discuss the creative and engineering design processes individually.

2.1.1 The Creative Process

Creativity research has long tried to synthesize its results through the elaboration of models of creativity. These models attempt to provide a common framework for further empirical research and aid in the design of creativity support tools (Shneiderman, 2000). One of the most influential early models of creativity was proposed by Wallas (1926). He divided the creative process into the four distinct phases of Preparation, Incubation, Illumination and Verification. Preparation involves gathering knowledge and understanding the problem. In the Incubation phase, the subconscious takes over, mulling over the problem without deliberate concentration. Illumination occurs as a sudden flash of light, when the solution has been discovered. Verification consists of evaluation of the newly formed idea.

From the four stage creative process model, a wealth of new models was born. For instance, Osborn (1963) broke the creative processes into two main phases of Idea Generation and Idea Evaluation, decreasing the emphasis on incubation. Amabile (1983b) developed a five stage model including Problem and Task Presentation, Preparation, Response Generation, Response Validation and Outcome. Shneiderman (2000) took a different approach to modeling creativity, classifying the types of creativity, and the creators, into the divisions of Inspirationalist,
Structuralist and Situationalist. He also developed a four stage model of Collect, Relate, Create and Donate; placing an emphasis on the contribution of creative act to the community (donate).

![Figure 1: Generic Creative Process Model adapted from Warr and O’Neill (2005).](image)

Warr and O’Neill (2005) synthesized the main creativity models into a unified model of Problem Preparation, Idea Generation and Idea Evaluation, see Figure 1. In the problem preparation phase, designers view relevant information associated with the problem in order to develop an understanding of what is required and to provide a basis for generating valid solutions. During the idea generation phase, designers develop new ideas which are then evaluated based on their appropriateness and creativeness in the evaluation phase. This Generic Creative Process model stresses the similarities of all previous models in an attempt to reach a uniform consensus.

Similar to Warr and O’Neill (2005), Howard, Culley and Dekonick (2008) performed a comparative summary of creative process models, but included more models in their analysis than Warr and O’Neill. They synthesized these models to contain the four stages of analysis, generation, evaluation and communication/implementation. However, they argued that the communication/implementation phase should be deemed a design activity and thus the generic
creative process model they developed contained the three stages of analysis, generation and evaluation. This model is strikingly similar to the model developed by Warr and O’Neill (2005). Thus, when referring to the creative process in this dissertation we refer to the stages of problem preparation (analysis), idea generation and idea evaluation.

2.1.2 *The Engineering Design Process*

Just like research in the creative design domain, there have been many attempts by researchers in engineering to develop models of the engineering design process. A well-known model of the engineering design process was developed by Booz et al. (1968) and included the linear processes of new product strategy development, idea generation, screening and evaluation, business analysis, development, testing and commercialization. Although the linear or sequential style presented in this model is by far the most dominant in the field, there are several other forms of process representations that have been published in engineering design including divergent-convergent models, knowledge space models and cyclic models, just to name a few.

Divergent-convergent engineering design process models differ from the linear models in that they assume some form of integrated evaluation and the selection of ideas and concepts. For example, Pugh (1991) created a model that included the stages of: market, specification, concept design, detail design, manufacture and sell. Pugh stated that the emergence of ideas is an interactive *cycle* that expands and contracts such that the best concepts emerge, and one has converged on to the concept to be selected, see Figure 2. In other words, one diverges and converges in the ideation process until the best concepts are identified.
Knowledge space models are a more atypical form of representation for the engineering design process. In these models, it is assumed that a certain amount of knowledge must be gained in each phase of the process before you can move to the next phase. For example, the C-K theory, or concept-knowledge theory, describes design as a process of movement between a concept and knowledge space (Hatchuel & Weil, 2003). Jansson and Smith (1991) developed a similar model that involved movement between the configuration space, which is an imaginary space which contains the mental representations of configurations such as diagrams and sketches, and the concept space, which contains ideas, relationships or other abstractions. The main thrust of this model is that movement within the configuration space must be obtained by movement to the concept space and then movement back to the configuration space, see Figure 3. These models serve as both a design theory as well as a theory of reasoning in design.
Although there are many forms of engineering design process models, Khandi (2005) argues that the basic five-step process typically used in problem-solving should be used to describe the engineering design process as well because similar to problem solving, design problems are typically vaguely defined and have numerous correct answers. In addition, the process may require some backtracking and iterations, and are thus more cyclic than linear. Therefore, Khandi defines the engineering design process as defining the problem, gathering pertinent information, generating multiple solutions, analyzing and selecting a solution and testing and implementing the solution, see Figure 4. In a more concise but similar summary, Sheppard (2003) characterized engineering as *scoping, generating, evaluating* and *realizing* ideas.
2.1.3  **Comparison of Creative and Engineering Design Processes**

The previous sections have brought to light several similarities in the creative and engineering design processes. Figure 5 displays a summary of the models developed by synthesizing the existing literature base in both engineering and creative design. As can be inferred from this graphic, the need to generate and evaluate ideas is present in all three models and thus is a central component of both engineering and creative design. In addition, the need for information, its analysis and understanding is also significant in both domains. Therefore, in the remainder of this document we will refer to the ‘design process’ according to these similarities.

Because design researchers and educators have noted the importance of evaluating individuals design skill not just by their ability to develop quality solutions, but also by their ability to *successfully* perform each step of this process (McKenzie, 2002), it’s important that we outline specifically what transpires in each of these phases before we can discuss existing computational tools developed to support each of these phases of design.
2.2 Design Process-Analysis

During the analysis phase of the design process designers analyze both the design problem as well as similar, existing designs. This analysis of the problem is an important first step in the design process since design problems are often open-ended and ill-defined (Lawson, 2006). Since analyzing involves parsing something into its parts (L. W. Anderson et al., 2001), the analysis of the problem requires the designer to ‘parse’ the problem in order to define the problem and develop constraints. In addition, individuals ‘parse’ existing solutions by decomposing existing design examples and identifying interesting components that can be reused in their own design ideas during the generation phase (Nickerson, 1999; Ward, 1995; Weisburg, 1995).

Since individuals reuse existing example components in their own ideas, it is important that the designer is able to differentiate better from poorer performance in existing designs so that the re-used components are of a high quality. This ability to differentiate example quality is related to individuals design expertise; the more design knowledge one gains, the more able they are to identify both high quality examples and useful example components (N. Bonnardel & Marmèche, 2005). This analysis of examples is an important and necessary part of the design process as both anecdotal and historical accounts have shown that even the most creative ideas are often developed through minor extensions of existing concepts (Ward, Smith, & Finke, 1999).

2.3 Design Process- Idea Generation

Once the designer understands the problem and has developed enough domain specific knowledge (through the analysis of the problem and existing solutions) they can start generating ideas. Designers should generate a variety of conceptual solutions prior to evaluating and
selecting concepts for further refinement and development as generating several ideas increases the chances of finding better ideas (Shah, Vargas-Hernandez, & Smith, 2003). Generating ideas involves transforming, combining or adapting elements of existing designs and knowledge in order to generate new ideas (L. W. Anderson, et al., 2001; C. Eckert, et al., 2005). Therefore one’s ability to generate ideas relies not only on their ability to analyze existing solutions, but also one’s existing knowledge structures. In fact, Herman Herzberger (1991) once said that,

“Everything that is absorbed and registered in your mind adds to the collection of ideas stored in the memory: a sort of library that you can consult whenever a problem arises. So, essentially the more you have seen, experienced, and absorbed, the more points of reference you will have.”

This quotation brings to mind several interesting questions. First, how is knowledge stored and retrieved from memory to aid in ideation and second, is knowledge always useful in design, or are there instances when it can actually constrain thought? We’ll begin this discussion by understanding how information is retrieved from memory for use in ideation.

2.3.1 Memory Search in Idea Generation

When referring to memory retrieval in idea generation, researchers typically use the phrase memory search instead of memory retrieval, as new ideas cannot be directly retrieved from memory (Nijstad, Stroebe, & Lodewijkz, 2002). On the other hand, idea generation involves retrieval processes as new ideas cannot be developed without existing knowledge (T. M. Amabile, 1983a). Therefore, memory search is used to describe the dynamic nature and role of the cues used to probe relevant memories in the design process. These cues can be either self-generated or provided as an external stimulus in the form of design examples.
In order to explain the role of cues in idea generation, Nijstad and colleagues (2002) developed SIAM (Search for Ideas in Associative Memory), a two phase process which includes a knowledge activation stage followed by an idea production stage, see Figure 6. In the knowledge activation stage, individuals rely on cues, to trigger an image (memory). Once generated, the cues are used to probe memory and activate related images. The probability of retrieving an image is dependent on the association between the search cue and the image.

This process falls in line the spreading activation model of memory that organizes images according to the strength of association (Collins & Loftus, 1975). The associative links between the images in this model represents the degree of semantic relatedness between the two concepts; the shorter the length the stronger the association and the faster an individual is likely to relate the concepts, see Figure 7. For instance, if the cue used was the word red, the individual shown in Figure 7 would be more likely to retrieve an image about fire engines than an image about a street because of the individual’s strength of association between these concepts.
Once an image is retrieved from memory, features or components of the image are used to generate ideas in the idea production stage. These ideas are developed by generative processes such as by developing new associations (Mednick, 1962). These ideas are then used as memory cues to activate and retrieve more images. This model relates to the spreading activation model in that presenting a concept (i.e. fire engine) first activates strongly linked concepts (i.e. red), followed by slightly weaker associations (i.e. car). This activation continues to spread to the more distantly related links, including associates (i.e. street and flower) but becomes weaker as it continues to spread (Sternberg & Ben-Zeev, 2001). Therefore, the use of cues (both internal and external) to search memory allow for the possibility of triggering many different associations.

Images that are successively activated during an idea generation session are likely to be semantically related. Because of this relationship, the activation of these successive images often leads to a train of thought and a rapid accumulation of semantically related ideas. When this
train of thought no longer leads to a new idea (i.e. the memory search is unsuccessful), a new search cue must be developed which is often a time consuming and conscious process. Once developed, this new cue is then used to probe memory, which leads to the activation of additional images and the generation of ideas. This process continues until no new cues can be developed or the retrieval of images no longer leads to new ideas.

2.3.2 Prior Knowledge and Idea Generation

So, does prior knowledge always aid in the retrieval of relevant information and the development of new ideas or are there instances where it can actually constrain thought? Creative cognition is a research thread developed to understand the way people obtain, organize, process, store and use information during creative thinking (Cropley, 1999; Finke, T. B. Ward, & Smith, 1992). This line of research assumes the same cognitive functions and processes involved in noncreative thinking can be used to explain creative thinking as well (S. M. Smith, Gerkens, & Shah, 2006). One thread of creative cognition research, referred to as conceptual expansion, is devoted to understanding what constrains individuals from transforming or extending existing concepts to develop new ideas (Ward, Finke, & Smith, 1995). In other words, although the mapping from old (knowledge or existing design) to new (ideas) can facilitate progress, it can also limit an individual’s ability to ‘think outside the box’ or move beyond familiar concepts to develop something truly unique.

Figure 8: Non-Earth creatures developed by participants in Ward (1994).
As an illustration of conceptual expansion and its limitations, Ward (1994) conducted an experiment where participants were asked to imagine (and draw) animals that lived on a planet very different from Earth. The novel creatures developed by the participants were then assessed for the presence of familiar and characteristic animal properties such as arms, legs and bilateral symmetry to see how the participant’s imagination was structured by their knowledge of Earth animals. The results showed that the vast majority of the creatures created displayed many similarities to common Earth animals, see Figure 8. These results were replicated even when participants were encouraged to develop aliens that were ‘wildly different’ from Earth animals (Ward & Sifonis, 1997). These findings suggest that individual’s knowledge about the typical features of familiar categories structure their imagination and their ability to develop ideas (Ward, 1995).

The constraining effects of category exemplars are not limited to the development of imaginative animals; as this effect has also proven to generalize to different conceptual domains (Perttulla & Liikkanen, 2006; Ward, 1994), ages (Cacciari, Levorato, & Cicona, 1997) and ability groups (Woodward, 2004). Ward (1994) proposed that these structuring effects can be attributed to individuals picking the ‘path of least resistance’ where they first retrieve specific instances of known concepts and then project those instances to new ideas (i.e. we recall images and then use them as a starting point for ideation). In other words, one’s existing knowledge structures can have a negative impact on one’s ability to think creatively.

In addition to long-term memory structures, one’s ability to generate ideas can be influenced by recently activated knowledge, such as when examples are given during a design task (S. M. Smith, Ward, & Schumacher, 1993). These examples are often provided as a way to demonstrate useful past approaches to problem solving, or provide an idea of what is currently on the market. However, like prior knowledge, these examples can often have an inhibiting effect on ideation as
they can constrain one's thought processes. This effect, referred to as cognitive fixation, has been extensively studied in both the problem solving and design literature. The following section will outline these effects and discuss the impact of this fixation on the ideation process.

2.3.3 Fixation in Design

Jansson and Smith (Jansson & Smith, 1991) were the first to provide an experimental approach to studying fixation effects in design. They hypothesized that showing a designer a picture of a design example, or a known solution to the problem, prior to idea generation would cause a mental block and reduce access to other ways of solving the problem. In order to test this theory, they conducted a series of experiments utilizing three different design tasks and mechanical engineering designers, both senior level undergraduate students and practitioners.

For each of the design problems tested, two groups were formed; one which received the problem description and a pictorial example and one which received only the problem description. The examples used in this experiment were specifically chosen to include elements that were in conflict with the problem description. The designers were given one hour to produce as many sketches of solutions to the design problem as possible. The sketches developed by the participants were then compared to a set of features included in the pictorial examples in order to judge the similarities of the sketches to the examples. It was expected that if fixation resulted from being exposed to pictorial examples, that more features of the examples would appear in the sketches developed by the designers in the ‘example group’. This was in fact found to be true; for both levels of expertise (students and practitioners) and across all design problems, more features associated with the example, including those that were deemed inappropriate, were found to appear in the sketches of the ‘example’ group. In other words,
designers that were shown an example demonstrated a lack of flexibility in the design process often reproducing the characteristics of the provided example.

Although Jansson and Smith’s (1991) study provided a foundational understanding of design fixation, it also seemed to generate more questions than it answered. For instance, is this design fixation effect restricted to engineering designers and does one’s familiarity with the elements in the example effect their fixation. Purcell and Gero (1996; 1993; 1992) sought to answer these questions through a series of experiments over a number of years. They followed the same methods of the previous experiment (Jansson & Smith, 1991), but they used novice design students in industrial design in addition to mechanical engineering.

The results from these experiments suggest that fixation, as described by Jansson and Smith, may only occur when designers lack domain specific knowledge (novice designers) and are forced to rely on everyday knowledge activated by a picture of a familiar example (A. R. Purcell & Gero, 1996). They also suggest that the complexity and originality of the design example may impose attentional constraints on the individual causing the designer to rely more on the elements in the example to provide a solution (A. R. Purcell, et al., 1993). In addition, mechanical engineers were shown to become fixated on features of the example when the examples embodied typical principles of the mechanical engineering knowledge base. On the contrary, when shown an innovative example, mechanical engineers became fixated (in a positive light) on the principles involved in the example, allowing them to explore new ways of solving the problem using this innovative principle. A follow up study by Perttula and Sipila (2007a) reported similar findings; presenting designers with common examples leads to a higher fixation effects than presenting designers with novel ones, see Figure 9 for an example. These findings suggest that the types of examples (based on familiarity and originality) utilized can have a significant impact on the type of fixation (positive or negative) encountered.
The evidence that fixation does occur in design practice is quite compelling. As a result, researchers generally acknowledge design fixation as a significant impediment on idea generation, thereby limiting the diversity and originality of creative outputs (Perttula & Sipila, 2007b). These previous studies discuss how design fixation is connected to designer’s expertise, their familiarity with the examples used, and how similar the examples are to the design task. However, no study to date has explored the types of examples designer collected on their own for a particular task, as subjects were specifically given examples to be used for inspiration in previous studies. Therefore, we are unaware if designers are collecting fixating examples in practice. If we can understand if or when designers collect fixating examples during a design task, and the search behaviors associated with this fixation, we may be able to develop tools to mitigate these effects and aid in the education of future designers.

Now that we have an understanding of how internal knowledge and existing example designs can aid and constrain the idea generation process, we can begin discussion on how designers evaluate the multiple ideas developed through these processes.
2.4 **Design Process- Evaluation**

Evaluation involves judging concepts and determining their value (L. W. Anderson, et al., 2001). Therefore, during the idea evaluation stage of the design process, the novel ideas produced during generation are evaluated based on their value in terms of their ‘appropriateness’ for the current problem and their novelty (A.F. Osborn, 1963). The evaluation of ideas is a vital part of the design process as finding and selecting proper ideas for further development has a definite influence on the quality of the final product (Bailey, 2006). Since a multitude of ideas are developed during the generation process, it is often difficult to develop criterion for selecting the best ideas. Because of this difficulty, a number of techniques have been developed to aid in the evaluation process at both the individual and group level.

At the individual level, designers can engage in a reflective evaluation in which the designer rejects ideas which do not solve the problem, are irrelevant to the current problem, or are clearly impossible to implement. The designer may also elicit opinions from experts about the ideas in order to identify potential problems with products or services before more comprehensive evaluations (J. Nielsen, 1994; Poulson, Ashby, & Richardson, 1996). Finally, one of the most popular evaluation techniques is the design critique or design review. Critiques can be collaborative such as receiving a design critique from a peer or colleague or an individual critiquing their own ideas (either systematically or intrinsically) (Herring, Jones, & Bailey, 2009). What differentiates critiques from other evaluation techniques is the *feedback* that results from the critique.

During the critique process, the designer presents their peers, colleagues and/or experts with their ideas, who evaluate them and provide feedback on the quality and originality of the ideas. In other words, critiques serve as not only a way to evaluate ideas, but also as a means to promote design learning and improve design ideas. Critiques can occur orally by the creator.
presenting their work to a jury of peers, colleagues and/or experts, or they can occur in a more intimate setting (such as online) where the creator sends the ideas to their peers for written feedback. Although the oral presentation is the traditional method of critique, an empirical study by Dannel's (2005) found that online settings are ideal for critiques in education because it minimizes power differences between critics and students.

In addition, although critiques are often preferred over other evaluation techniques because of the feedback provided, the quality of the feedback has a large influence on the utility of the critiques and its ability to help in the evaluation of ideas. In fact, research has found that critiques of ideas that contain meaning-level feedback have been shown to lead to substantial student revisions that positively affect the ideas whereas rubber-stamp feedback, or feedback that focuses primarily on errors in the design, do not result in the same positive effect (Dannels, 2005). In addition, researchers have found that overly harsh and direct feedback may slow idea improvements whereas motivating feedback improves performance (C. D. Smith & King, 2004).

Finally, studies have shown that the quality of an evaluation/ critique is heavily reliant on domain-relevant skills, as the individual must have knowledge by which to assess the appropriateness of the generated ideas (T. M. Amabile, 1983a). Therefore, individuals with more design knowledge often provide more useful design critiques. These results show that the value of the critique depends on the quality of the feedback received which can be influenced by who provides the critique.

2.5 **Computational Tools for Supporting Idea Generation in Design**

One way researchers have tried to help designers in the design process is by developing computational tools that support human cognition and design skill development. In fact, the
aspiration of most creative design research is to inform the creation of tools to improve the efficiency of the design process and the quality of ideas developed (Shneiderman, 2000). A number of researchers have proposed standards and implications for the design of creativity support tools and have suggested implications for the design of creative support systems focusing on fine arts collaboration (Thomas, 2002), sketching (Gross & Yi-Luen Do, 1996), creative problem solving environments (Hewett, 2005), distributed scientific communities (Farooq, 2005) and a host of other fields.

In fact, a number of tools have been constructed with the goal of improving the creative process. The Electronic Paper Napkin helps designers retain the ambiguity in their designs and attempts to intelligently recognize what is being drawn (Gross & Yi-Luen Do, 1996). Another physical based idea generation tool is SILK, an interactive sketching tool designed to facilitate rapid prototyping. This tool aids designers in building rapid prototypes of interfaces through a unique sketching interface (Landay & Myers, 1995). Tools such as IdeaTree and IdeaFisher provide the user with associative linking; however a study of these tools found them insufficient for practical use (Massetti, 1996). While a wealth of these tools has been developed, they are not utilized within industry and thus more studies are needed that focus on the inefficiencies in current practices in design, which is the focus of Part I of this dissertation.

In addition to tools developed to improve the creative design process, there are also tools which have been developed to help in the collection of design examples that aid both in the development of new ideas as well as the triggering of relevant memories for design ideation. Although there are many sources that can be used to locate these sources (books, magazines, etc.), the Web is widely used for information finding and prior work has shown that designers perceive the Web as a rich source of examples (Herring, et al., 2009). Though not the specific focus, there has been research that could improve the Web for example finding. For instance, in
the area of image search, Yee et al. (2003) developed a category-based method that enables users to navigate along the conceptual dimensions of an image set. Fogarty et al. (2008) developed a new interface that allows users to perform image search through visual composition of search rules. Cuil, a new search website, renders images to illustrate the main concepts behind search results and arranges the results in a 2D space (“Cuil - Our Philosophy,”). In addition, case-based systems have been grounded in the perspective that designers solve new problems by relating them to prior experiences (Schank & Leake, 1989). Research on these systems has focused on developing computational representations of cases and algorithms for efficient retrieval and has been studied mostly in the engineering design domains.

These and other improvements to search interfaces have the potential to aid designers in example finding and thus the first two phases of the design process (preparation/analysis and idea generation). However, because less is known about how designers use the Web for example retrieval and what types of examples they collect it is not clear whether these techniques would help for this search context. Consistent with findings reported by Pearce et al (1992), this area of research would benefit from deeper understanding of how designers find and select examples in practice and how tools can better support example finding behavior.

2.6 Summary of Engineering Design Literature

This chapter has discussed the engineering design process and it’s relation to the creative process and the way individuals use external and internal sources to develop ideas. In addition, it highlighted the opportunity for the development of computational tools that better support the design process and the reuse of information during ideation.
Part I of this dissertation is devoted to understand the ideation strategies utilized in professional practice to develop ideas and in particular the use of examples during this process to inform the development of computational tools. Therefore, the following three chapters will outline three studies developed to identify opportunities for design tools that better support these practices and highlight a new tool, TweetSpiration, which was developed to address some of these challenges. Following these three chapters, we will focus on how design skill is currently assessed during each of these three phases of design in engineering education, and how examples may be used to better assess individual design competence, which is the focus of Part II of this dissertation.
Chapter 3

Experiment 1: Idea Generation Techniques

The goal of the first three studies in this dissertation is to outline the ideation practices currently used in professional design practice and highlight the utility of example-based methods that support ideation. Through this knowledge we can develop implications for new design tools that can better assist designers in these processes. In the current chapter, we introduce a study developed to understand designer’s perception of the role of creativity in the design process and the idea generation techniques used in professional practice. This is an important study because although some research has focused on developing technologies to better support the cognitive processes involved in creativity, much of the research in the field has been focused on highly formalized idea generation practices during group design sessions, without placing these sessions within the larger context of design (Lin, Hong, Hwang, & Ling, 2006; Paulus & Yang, 2000; Sutton & Hargadon, 1996; Van der Lught, 2005). Without the foundational understanding of the creative process and the idea generation techniques utilized in practice, it is difficult to develop tools that will support and enhance the design process.

Now, there have been numerous research studies that attempt to provide a common framework for empirical research and aid in the design of creativity support tools, but there is some skepticism regarding the simplicity of these current models, see section 2.1. For example, most
of these models note that they are not intended to be followed in a discrete linear fashion (A. Warr & O'Neill, 2005), while the representations are always portrayed in a static, linear fashion cycling through distinct stages of the creative process, see Figure 1. Many authors have rejected this distinct and limited representation (Eindhoven, 1952; Ghiselin, 1963; Lubart, 2001); arguing that creativity is a “dynamic blend of processes that co-occur, in a recursive way throughout the work” (Lubart, 2001). These same authors call for a better understanding of the creative process, and a better representative model. This study seeks to fill this gap in knowledge by providing a new model of the creative process based on interviews with professional designers.

In addition to providing a new model of creative design, it is also important that we understand the ideation techniques utilized in professional practice if we intend to develop guidelines for future design tools. For example, while there have been over 172 different idea generation techniques identified in the literature including, but are not limited to, Osborn's (A.F. Osborn, 1963) Brainstorming and SCAMPER (substitute, combine, adapt, modify, put to other use, eliminate, rearrange) (G. J. Smith, 1998), there are a limited number of studies that have addressed the frequency of use of idea generation techniques and their applicability during constrained situations (Lin, et al., 2006). Knowledge of the relative importance of these techniques is crucial for informing tool design, as only a few of the techniques are used frequently in practical design situations. Thus the purpose of this study was to provide the foundational understanding of the role of creativity in design and the benefits and limitations of current idea generation practices in order to develop guidelines for computer based tools that truly support these processes.
3.1 **Methodology**

The first study in this dissertation was developed to understand the design process as viewed from the design community, develop a framework for the idea generation process, gain insights on current strategies used to generate ideas and establish a develop guidelines for future design tools. This section describes the participants, the interview script developed, metrics used for comparison, the experimental design and the results.

3.1.1 **Participants**

Ten professional designers involved in the product design process were interviewed in this study. Nine of these designers were from three different companies and one was a freelance designer. Seven of the interviews were conducted in the designers’ project space, which facilitated the observations of artifacts and allowed us to observe their working environment while the other three interviews were conducted over the phone due to geographical constraints.

<table>
<thead>
<tr>
<th>Professional Background</th>
<th># of participants</th>
</tr>
</thead>
<tbody>
<tr>
<td>Engineer</td>
<td>5</td>
</tr>
<tr>
<td>Industrial Design</td>
<td>4</td>
</tr>
<tr>
<td>Product Design</td>
<td>3</td>
</tr>
<tr>
<td>Occupational Health</td>
<td>1</td>
</tr>
<tr>
<td>Architectural design</td>
<td>1</td>
</tr>
<tr>
<td>Graphic design</td>
<td>1</td>
</tr>
</tbody>
</table>

*Table 1: Designers' Professional Experience from Experiment 1.*

Two of the three companies we observed were large corporations with internal product development groups. These two companies were situated in the service and electronics industries. The other company we observed was a design firm that is typically contracted by outside clients to design end consumer products. The designers interviewed represented a range of professional design experience levels (Table 1) and came from diverse backgrounds (Table 2).
<table>
<thead>
<tr>
<th>Years of Experience</th>
<th># of participants</th>
</tr>
</thead>
<tbody>
<tr>
<td>Less than 5 years</td>
<td>2</td>
</tr>
<tr>
<td>Between 5 and 10</td>
<td>3</td>
</tr>
<tr>
<td>Between 10 and 15</td>
<td>3</td>
</tr>
<tr>
<td>Between 15 and 20</td>
<td>1</td>
</tr>
<tr>
<td>More than 20</td>
<td>1</td>
</tr>
</tbody>
</table>

Table 2: Designers’ Background from Experiment 1.

Some of the designers had professional background in many industries such as product design and occupational health or engineering and industrial design. Due to this, the background of the participants in Table 1 does not add up to the number of designers that actually participated in this study (10).

Figure 10: List of interview questions asked in Experiment 1.
3.1.2 **Procedure**

Upon meeting with the experimenter, the participant was asked to read and sign an informed consent document. Any questions were answered at this time. Once completed, an audio recorder was started and the semi-structured interview commenced, see Figure 10 for a list of interview questions. The interview lasted about 90 minutes.

3.1.3 **A note on observations**

The remainder of this chapter presents the observations of product design practice based on the current study. Due to the diverse background of our participants, much of what was observed during this study is not necessarily unique to product design but probably draws from the broader traditions of design including (but not necessarily limited to): architectural design, industrial design, graphic design and engineering design. The purpose of this study was not to determine what was unique to product design but rather to learn about current practices in the field to guide and develop tools to better support those practices.

3.2 **Results**

This section serves as a summary of the results from Experiment 1.

3.2.1 **Design cycle model**

When participants were asked to describe their creative process, the designers described their entire product development cycle from idea creation to client presentation. The designers viewed their overall design process as a creative process, referencing how it is important to be creative throughout the entire development cycle.
Previous models of creativity focus mainly on the idea generation process and do not differentiate ideas from final product solutions, see Figure 1 as an example. Following the responses from designers, the Design Cycle Model was developed which makes the distinction between the design cycle and the idea generation process, see Figure 11.

The design cycle consists of three phases: *idea generation*, *implementation* and *evaluation*. These phases are represented as distinct circles due to the deliberate separation of the design cycle by time management practices. Under a strict deadline, designers must ensure that they place clear boundaries between the phases in the cycle, ensuring that the product is delivered to the client on schedule. When the designer leaves the idea generation circle he or she has an idea that she is ready to implement. Following the completion of the solution, the designer evaluates the solution and its representation. This evaluation leads to a new insights utilized in the next iteration of the design cycle.
3.2.2 \textit{IR^3 Idea Generation Model}

Focusing on the idea generation phase we created the IR^3 Idea Generation Model, see Figure 12. This model describes the fluid cycle of idea generation utilized by designers during the conceptual stage of the design cycle. Within this model the designer generates and refines ideas, eventually leaving the circle with an idea that is ready to implement and moving to the next cycle in the overarching design cycle.

The model consists of three non-distinct categories: research, represent and refine all encompassed by the category of inspiration. Designers search for inspiration throughout the idea generation process in order to spark the formation of creative solutions. The process begins with research into the problem domain which will generate a multitude of diverging concepts. Once initial concepts are formed in the research phase they are physically represented in an externalized form. When the representations are solidified the designer begins to evaluate and refine the concepts eventually leading to a convergence of concepts. The refined, validated concepts then guide further research and knowledge acquisition starting the process anew. The cycle by no means maintains a constant speed of rotation. During the idea generation process a concept may be represented by a quick sketch or may be discarded during a split second refine cycle. (A detailed explanation of each stage occurs in the Categorization of Techniques section.)
This model depicts the transference of concepts and not ideas. This subtle distinction is made to explain the initial problem finding exploration, within which the designers attempt to define a *wicked* problem. Thus a concept represents either a working definition of the problem or a potential solution.

### 3.2.3 Idea generation techniques

The audio recordings of the interviews were analyzed by the authors and every reference to a distinct idea generation technique was documented. Then similar idea generation techniques were conglomerated into larger categories. For example all references to emailing, instant messaging or asking someone a specific question were grouped into a *consultation* category. These categories were carefully refined until agreement was met between the authors.

According to Smith (1998) 172 idea generation techniques have been identified. According to the results of our interviews, 19 techniques were mentioned as appropriate idea generation strategies for design. The idea generation techniques identified are briefly introduced as follows:

1. **Role Playing**: Role playing involves designers acting out scenarios. These scenarios are often ones that the designers observed during the research phase of the design process when they participated in user research. This technique is used as a tool for both team-based ideation and communication to users and/or clients (Alderson-Williams, Bound, & Coleman, 1999; Moore & Conn, 1985).
2. **Active Search**: Active search refers to designers hunting for a particular solution. This hunt could range from a web search for images of current vacuum cleaners to searching through books, magazines, newspapers, etc. to find the demographics of a particular population (Lee, Klemmer, Srivastava, & Brafman, 2007).
3. **Attribute List**: Attribute listing refers to taking an existing product or system, breaking it into parts and then recombining these to identify new forms of the product or system (Crawford, *The Techniques of Creative Thinking*; Morgan, 1993).
4. **Brainstorm**: Brainstorming involves generating a large number of ideas or solutions to a problem with a focus on quantity of ideas. During this process, no ideas are evaluated; in fact unusual ideas are welcomed. Ideas are often combined to form a single very good idea as suggested by the slogan “1+1=3” (A.F. Osborn, 1963). Brainstorming can be used by groups as well as individuals (Furnham & Yazdanpanahi, 1995). Since brainstorming was
the first idea generation technique created it is often referred to as, “the mother of all idea generation techniques” (A. F. Osborn, Rona, Dupont, & Armand, 1971).

5. Collaborate: Collaboration refers to two or more people working together towards a common goal (Myrian-Webster's - Online Dictionary, 2007). Designers often work in groups and co-create during the entire creative process.

6. Concrete Stimuli: Concrete stimuli are used when designers want to gain new perspectives on a problem by manipulating physical materials. This could be looking at paint chips, feeling different material textures or physically maneuvering objects.

7. Critique: Critique refers to receiving input on your current design ideas. This could be collaborative such as receiving a design critique from a colleague or individuals critiquing their own ideas (either systematically or intrinsically). This technique often spurs new thought by finding solutions to design flaws within current concepts.

8. Documenting: Documenting refers to designers writing down ideas (physically or electronically). This includes journaling, writing stories, and taking notes.

9. Expert Opinion: Designers often elicit opinions from experts to identify potential problems with products or services before more comprehensive evaluations. This occurs when they are looking for an answer to a problem that is outside their domain knowledge or when they want to test a new idea (J Nielsen, 1994; Poulson, et al., 1996).

10. Empathy/User Research: User research requires the designer to observe real people in everyday situations in order to develop empathy for the users. Many of the methods used to conduct this type of research are founded in ethnographic research methods such as observations, field studies and rapid ethnography (Aldsersey-Williams, et al., 1999; Norman, 1998).

11. Encompass: Encompassing is an inspirational technique which involves designers immersing themselves in information relevant to the current project.

12. Forced Analogy: Forced analogy involves comparing the current problem with something else that has little or nothing in common in order to gain new insights and results. This technique often generates ideas for new areas of research. (Huhns & Acosta, 1988; McAdams & Wood, 2002)

13. Incubate: Incubation refers to stepping back from the problem to let the subconscious mind work (Wallas, 1926).

14. Passive Searching: Passive searching refers to when designers look through material (web, magazines, books) for inspiration; they are not searching for a particular solution to a problem, they are merely looking to be inspired.

15. Prototyping: Prototyping, in this study, refers to a low-fidelity model of an idea. These models can be created with any type of material (paper, clay, etc.) as they are only used to conceptualize a thought.

16. Reflect: Reflection occurs when designers review their previous work (sketches, documents, prototypes, etc.)

17. Sketching: Sketching refers to a rough drawing of an idea.

18. Socializing: Socializing refers to talking with others about topics unrelated to the current project.

19. Storyboards: Storyboards are a way for designers to represent information gained in the research phase of the design process. Quotes from the user, pictures, and other relative information are placed on cork board, or a similar surface, to represent a scenario and to help understand the relationships between design ideas. Designers often post information about users using as little detail as possible to allow for interpretation of information (Beyer & KHoltzblatt, 1999; Newman & Landay, 2000).
Categorization of techniques

After the 19 idea generation techniques were identified, they were categorized and placed into the IR^3 model according to what the technique was used for (inspiration, research, representation or refinement). The techniques were placed into the boundary areas of the model if the technique served many purposes. Therefore, some techniques are placed on the boundaries between two categories or in the middle of the model if they encompass all aspects of the ideation process, see Table 3. A detailed description of the categories and the explanations of technique placement are described below.

<table>
<thead>
<tr>
<th>Technique (% Referenced)</th>
<th>Research</th>
<th>Represent</th>
<th>Refine</th>
<th>Inspire</th>
</tr>
</thead>
<tbody>
<tr>
<td>Active Search (100)</td>
<td>x</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Attribute List (40)</td>
<td></td>
<td>x</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Brainstorm (80)</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td></td>
</tr>
<tr>
<td>Collaborate (60)</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
</tr>
<tr>
<td>Concrete Stimuli (2)</td>
<td>x</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Critique (90)</td>
<td></td>
<td></td>
<td>x</td>
<td>x</td>
</tr>
<tr>
<td>Documenting (60)</td>
<td>x</td>
<td>x</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Expert Opinion (90)</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td></td>
</tr>
<tr>
<td>Empathy/ User Research (80)</td>
<td>x</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Encompass (50)</td>
<td></td>
<td></td>
<td></td>
<td>x</td>
</tr>
<tr>
<td>Forced Analogy (5)</td>
<td>x</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Incubate (30)</td>
<td></td>
<td></td>
<td></td>
<td>x</td>
</tr>
<tr>
<td>Passive Searching (60)</td>
<td></td>
<td></td>
<td></td>
<td>x</td>
</tr>
<tr>
<td>Prototyping (70)</td>
<td>x</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Reflect (60)</td>
<td>x</td>
<td>x</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Role Playing (10)</td>
<td></td>
<td></td>
<td></td>
<td>x</td>
</tr>
<tr>
<td>Sketching (100)</td>
<td>x</td>
<td>x</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Socialize (30)</td>
<td></td>
<td></td>
<td></td>
<td>x</td>
</tr>
<tr>
<td>Storyboarding (40)</td>
<td>x</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table 3: Categorization of idea generation techniques
Inspiration

As previously mentioned, inspiration can occur at any stage within the idea generation process. Many designers mentioned solely inspirational idea generation techniques such as: passive searching, encompassing themselves in the material, socializing, and incubating. Passive searching refers to searching without a particular concept in mind. In particular they search to, “Get inspiration, (when you) don’t even know what you are looking for”. Many designers reported searching through magazines, web sites, and books. Designers reported checking web blogs on a daily basis. Many designers also said they frequently search on the internet, “The internet is obviously a god-send. There are a few good websites that people frequent around here. (We) Spend time just looking through ideas.”

Designers also reported immersing themselves in information relevant to the current project, “It helps me to be really physical. You took photos, you took video. Maybe you built some prototypes. But you try to get it up in your space and spread it all out.” Designers reported going into their project space and reflecting, looking at all of the information and “finding out what is really important”.

Another method designer’s use is to socialize. Socializing refers to talking to someone about anything unrelated to the project at hand such as the weather, sports or politics. This allows them to clear their mind for new ideas to immerge.

Finally, designers incubate in order to find inspiration or let the information “sink in”. Some of the incubation methods mentioned by designers were taking showers and going on walks.
Research

During the research phase of the idea generation process, designers try to gain additional knowledge to help them identify potential solutions. This acquisition of knowledge could include anything from user research, to active web searches, to reflecting on previous work.

In this experiment, designers reported the necessity to have up front empathy with the users. In order for designs to achieve empathy with their users, they must conduct user research. Although designers often have a clear audience in mind, they often use forced analogies to develop a larger research area. For example, one designer was developing a new interior cab design for a commercial truck. In order to develop empathy for the user, the designer interviewed people who lived in small spaces such as small houses and semi sleepers. This forced analogy allows designers to research a broader area and gain additional knowledge.

In addition to the upfront user research, designers often reflect on their previous work. This work often comes in the form of pictures, physical prototypes, reports, sketches, etc. Designers can also perform an active search where they search for particular information. For example, if a designer is designing a new water bottle they could perform a web search for relevant pictures of water bottles or names of manufacturing companies. Although this type of search could take place in the web, it could also be performed by looking through books, magazines, newspapers, etc.

Sketches or documentation (physical or electronic) is also utilized during the research phase. The designer sketches possible solutions (ideas) as well as documents his or her journey by journaling (writing stories/ notes), “Sometimes the best way for me to think is to sketch... to sketch and to write.... 90% of my notebooks is stuff I’m thinking... (it helps me) organize my thoughts.” Designers often use these sketches and document to generate ideas in the next phase.
“Sometimes I get to the best sketch by writing, sometimes I get to the best story by sketching”.

**Represent**

During this represent phase of the idea generation process, designers use physical objects to represent their design ideas. As previously stated, *sketching* and *documenting* are both used during the representation phase of the idea generation process. However, in this phase the designers use sketching and documenting as a way to conceptualize their idea, “It helps other people understand your idea. It gives them something to react to”. Another designer stated that they liked to communicate their ideas through *acting* them out. This allows designers to properly communicate an idea to others when they cannot explain it by sketching and writing. *Word lists* also provide a way for designers to illustrate their ideas. When using word lists, designers list important features that should be included in the design. Word lists can include aesthetic as well as functional elements of a product.

*Storyboards* are also used by designers as a way of displaying knowledge gained during the research phase. Storyboards usually include information about users researched in the research phase. The designers represent these individuals by including pictures and a few key pieces of information that they learned during their visit.

“Quick and dirty” *prototypes* are also created in this phase to manifest their ideas by using materials such as foam core, paper, Legos™, and Play-Doh™. Prototyping in the representation phase should not be confused with prototyping in the later phases of the design process as it is only used to represent a rough idea and not a finalized product (low versus high fidelity prototypes).
When the designer needs to develop a product that is outside their area of expertise they often *elicit expertise* from others. The designers usually either search for an answer to the problem or a “spark that helps them form the answer”.

**Refine**

The final phase in the idea generation process is the refine phase. In this phase, designers converge ideas based on evaluation of the concepts. This evaluation could be a formal process through such as a design *critiques* from another individuals, or they could be informal critique from the individual such as having the notion that, “this idea won’t work”. The number of ideas is normally reduced in this phase as they weed out ideas, but by critiquing these alternatives they gain new insight. Additionally, designers often *reflect* on previous designs, and *actively search* for comparisons during critiques.

**Co-Creation**

There are two idea generation techniques that have yet to be discussed, collaboration and brainstorming. These two techniques appear in the middle of Figure 12 because they occur in all areas of the idea generation process: research, representation and refine. Collaboration refers to working with another individual to actively generate ideas. One designer described collaboration as, “sitting down and talking about it, seeing what seems to resonate, creating a loose construction of ideas in an informal setting.” This could occur during research by trying to identify possible research paths, discussing the ideas generated during the representation phase or by reflecting on the concepts created with another person.

*Brainstorming* is one of the most popular techniques used by designers. This is directly reflected by the 80 percent of the designers in our study who mentioned brainstorming as an
idea generation technique. In a brainstorming session, designers create as many ideas as possible in about an hour session. Within these sessions the individuals go through many cycles of the idea generation process by researching what others have presented, representing their own ideas, intrinsically critiquing the ideas in order to develop new ideas.

3.2.5 *Utilization of Idea Generation Techniques*

The percentage of times each technique was referenced was calculated to determine how many designers actually used the techniques. This calculation was completed by summing the total number of subjects that referenced the technique over the total number of subjects. Additionally, the percentage of techniques that fell into each category was generated, see Figure 13. This percentage was calculated by summing the total number of subjects that referenced the techniques in each category and dividing it by the total number of techniques mentioned by all subjects (results are truncated).

![Figure 13: Percentage of idea generation techniques referenced in each phase](image-url)
3.3 Summary and Discussion

This goal of this study was to understand the creative process of professional designers in order to develop better creativity support tools. We realized the importance of creativity throughout the design cycle and reflected this in a new creative model of design. Additionally, in order to deepen the understanding of the idea generation process, the IR³ model of idea generation was developed, emphasizing the importance of non-distinct phases and the cycle within a cycle concept. We also expanded the design literature base by providing new insights on the idea generation techniques used by designers in professional practice.

The next goal of this study was to take the lessons learned from the interview data and translate them into implications for tools to better support the design process. In order to develop these guidelines, we used the data from our participants where they discussed not only the techniques they used to generate ideas during the design process, but also the limitations of these techniques. The rest of this section will discuss these limitations and offer implications for future design tools.

3.3.1 Support Group Collaboration

“A lot of what we do is co-creation, (we are) building things together”

Most of the designers interviewed worked in a corporate setting, so they were almost always co-located with their affiliates. This is often useful as they critique colleagues design ideas, collaborate on products and socialize to inspire new thoughts. Although these things are easy to do when people are in the same location, it is difficult to replicate this type of community when people are distributed. Designers mentioned struggling to communicate design ideas when they
were working with team members that were dispersed throughout the city, state, country and world. Although briefly mentioning collaborative tools such as wiki’s and video conferencing, they immediately dismissed these technologies because their needs were never met. For example, one designer complained that sometimes files would be saved over on a wiki because no one knew who had the most up-to-date file or who made the last revisions. They also complained that video conferencing was inefficient, they liked the privacy of telephones (people couldn’t see what they were doing).

Due to the technology-designer mismatch, designers often revert back to old technologies such as emailing and telephones because they feel they are more effective at communicating and sharing ideas through these metrics. Software is needed that enables people that are distributed to have the same resources as those that are co-located (automated design critique, built in social channels, and collaboration tools that allow easy exchange of both ideas and file sharing).

3.3.2 *Elicit expertise*

Almost all of the designers that were interviewed mentioned eliciting expertise at some point during the design process. They claimed to utilize this technique when the problem was outside of their domain knowledge, when they were looking for a spark for an answer, and when they wanted to test a new idea. Although designers established the necessity of eliciting expertise, they outlined the difficulties of finding the right person to contact. One designer stated, "we have internet tools that are internal, you can go to people pages and see things about a person and read up on them, but in my experience the most useful thing is just word of mouth." This shows that there is some added benefits (both trust in the system (person) and ease of
information access (who to contact)) of addressing individuals you know as opposed to using these online information pages.

Although in most cases asking a colleague first hand is convenient, it is often difficult as designers work odd hours and are constantly traveling. When one designer was asked how he/she determines who to contact when expertise is needed they replied, “more often than not the first thing that you want to do is just email the whole office, or the whole company. You say, ‘Hey has anyone worked on a project about open carousels or on a project that has to deal with injectable devices.’ You want to throw that out there so you can get names of people in the company to talk to them face-to-face.” When another designer was asked how they elicit expertise they stated, “I don't like sending out general emails because there are so many general emails and there is a lot of time wasted so if I can avoid it, I do.”

Software needs to be developed that aids designers in finding the expertise they need for a particular design problem. Future research should focus on understanding the problems with current online, internal, tools and the information cues used by contacting a colleague face-to-face (how they know the expert, what projects they have worked on together in the past, etc.).

3.3.3 Re-use design knowledge

During the interview process, 60% of the designers queried mentioned reflecting on their previous work. Since all of the designers mentioned sketching, 60% mentioned documenting, 80% mentioned user research (including taking photos, videos and notes) and 70% mentioned making low fidelity prototypes, the need for a proper storage and retrieval system of design knowledge, at both the individual and group level, is greatly needed. Designers often cited the poor structure of their current electronic storage devices. At the individual level, designers need
to have a proper filing system that allows them to reflect on previous work without any limitations. Although this is already a challenge at the individual level (creating key words, making information readily accessible) it becomes even harder at a group level.

Designers often mentioned titling files with ambiguous names such as CellPhoneModel2.sldprt. The problem with this type of filing system is that it makes file recognition from other colleagues incredible difficult. In addition, designers not only want to have design knowledge of ideas generated, they want to know about deliverables and about company expectations. When asked about what types of information they reflect on during design, one designer stated, “In the early stage it could be trying to figure out the right avenues to do research in. In the later stages it is more important to learn about deliverables and how they can be communicated. (For instance) what tools were used... what obstacles had to be overcome...”

New technologies are needed to aid in design knowledge reuse. The system should be easy to implement (aka not having to scan in all sketches) and should be easy to search. Not only should generated ideas be stored, but also deliverables. The information should have several keyword search mechanisms such as by company, product designed, colleagues who worked on the project, etc.

3.3.4 Supporting active search

There were two types of search tasks mentioned during the interviews: active and passive. Designers mentioned using active search when they were looking for a particular solution. However, one designer summed up active search problems best when they said, “I find the internet really hard to use from a design standpoint. (AKA) Find the answer to our design problem through Google™.” For example, “If what you are after is not necessarily about
bathroom soap products, it is hard to find the keywords to define the projects that are relevant for the content, for the types of deliverables. You never really find easily what you are looking for. You have to already know what to look for.”

There are two types of problems associated with active searching: the user cannot find the information or the information is simply not available. Not being able to find the information is a problem that can be solved by encouraging designers to participate in tagging activities. For example, designers mentioned that they frequently visited blogging sites. If the designers found a blog that would be of particular interest for someone working on a new cleaning product, they could tag the blog with the appropriate keywords. This type of methods could be used not only on blogs but websites, pictures, magazines, books and other documents. Although seemingly simple, this type of group tagging activity could greatly benefit the design community.

The second problem with active search is that the information is not available. Often times designers search for design solutions to a particular problem (what type of spring to use for a particular function) and are unable to find an answer. It is not necessarily the case that no one has ever tackled the problem, but that the information is not readily accessible. This type of problem could be addressed by feeding into the expertise search engine. If someone had worked on a similar problem they could offer new insights for a solution.

3.3.5 **Supporting passive inspiration**

Inspirational techniques are often times very insightful for designers. Half of the designers mentioned encompassing themselves in material that is related to the project such as pictures, notes, similar products etc. There are many benefits to these techniques. One designer stated
that it helps them “be really physical. You took photos, you took video. Maybe you built some prototype. But you try to get it up in your space and spread it all out.”

When working on a design team that is co-located, this activity is extremely easy; there is one project space where all the information is stored. On the individual level this is also easy, as long as the person is in the same location as the design material. However, if the designer leaves to travel or the group is dispersed, this project space is no longer useful.

Software that supports the replication of these project spaces is needed. A dynamic digital picture frame could be displayed in each of the dispersed groups work space as a means of replicating this concept. As designer members added new pictures to the design folder, the picture frame would automatically update. This would allow all of the design team members to see new pictures, notes and ideas easily.

3.4 Future Directions

Design research has been conducted for almost a century and yet new findings are still being reported. This study has attempted to understand the creative process of professional designers in order to develop better computer-based design tools. The results from this study outlined five implications for future tools that better support the design process and outlined the importance of actively and passively searching for information, especially design examples, as part of the ideation process. Although this study outlined the importance of examples in the design process, the details of how and why examples are used remain an enigma. This is problematic for both the understanding of design activity as well as for developing more effective design tools that support actively and passively searching for these artifacts. The following chapter explores the
role of examples in the design process and offers guidelines for more effective tools that support the retrieval, storage and dissemination of examples.
Chapter 4

Experiment 2: Example Usage in Design

As Experiment 1 demonstrated, actively and passively searching for artifacts (examples) is an important part of the design process and finding and sharing them is common in many design disciplines. For instance, Figure 14 shows a project workspace in which a design team has immersed themselves with myriad examples collected during the design process. The term *example* is used to mean any material, product, prototype, or digital artifact that contributes directly or indirectly to a design. This could be a sketch, a webpage or any item that inspires creative thought. Though example use is pervasive in design, this process has not been widely studied.

![Figure 14: A design project workspace filled with examples.](image)
However, as discussed in Chapter 2, there has been some research conducted to assess how the use of examples both similar to a design problem (N Bonnardel, 1999) and familiar to the designer (A. T. Purcell & Gero, 1992) affects the design process and outcomes. However, these studies have not investigated which examples designers themselves would choose and why, or the different roles that examples serve in practice.

In this chapter, we report the results of a field study that aims to understand and compare the practices of example usage throughout the design process within three design domains: Web, product, and graphic design. Our field study consisted of semi-structured interviews with eleven professional and highly-skilled designers. The study presented in this chapter builds on Experiment 1 where we found that searching the Web, magazines, and books was an important part of designers’ ideation practices. The current work now studies this enigmatic behavior in much greater detail.

### 4.1 Methodology

This section describes the participants, the interview script developed, metrics used for comparison, the experimental design and the results for Experiment 2.

#### 4.1.1 Participants

Participants were recruited via electronic postings on the IDSA-Chicago website and Coroflot, the portfolio website for the Core77 design blog. Eleven designers responded and were interviewed for this study. Four participants were freelance designers while the others worked for one of several design firms in a metropolitan area. The firms are well known in the design community and have a recognized reputation for design innovation. All participants were
typically contracted by external clients to design websites, physical products, visual logos, or product packaging. The experience of the participants ranged from 1 to 17 years with an average of about 9 years. The participants had diverse educational backgrounds including a BFA in photography and a MFA in Human-Centered Communication design.

During the interview, participants were asked to describe their professional background for classification into product, graphic or Web design. Most of the designers had professional backgrounds in multiple design domains and were therefore classified into multiple categories. The classifications were 7 product, 4 graphic, and 3 Web designers. By interviewing a diverse set of creative designers, we were able to identify similarities and differences of example usage among the different domains.

4.1.2 *Procedure*

Upon meeting with the experimenter, the participant was asked to read and sign an informed consent document. Any questions were answered at this time. Once completed, an audio recorder was started and the interview commenced. The interview script consisted of 16 questions covering the benefits and limitations of examples; example retrieval, storage and dissemination techniques; and the designers’ attitudes toward contributing personal examples, see Figure 15 for a complete list of the interview questions. The semi-structured interview lasted approximately 90 minutes and conducted in a designer’s own workspace to allow the designers access to their design materials and allow us to observe their work environment. A digital camera was used to collect images of interesting artifacts or observations during the interview.
Background
1. Please describe your educational background.
2. What is your professional background?
3. How many years of industry experience have you attained (please describe)?
4. Do you consider yourself an expert (briefly explain)?

Benefits of Examples
5. Can you briefly describe a recent or ongoing project where you used examples in the design process?
   a. Can you show some of the examples that were used?
   b. How did the use of these examples benefit you, if at all?
   c. How did you find these examples?
6. For what purposes do you typically utilize examples during the design process (similar products, photographs, color palettes, etc.) (Can you show us examples of use)?
7. What types of examples are most useful?
8. In which stages of the design process are examples most useful and why?
9. For what reasons do you consciously choose not to look at examples, if at all?

Example Retrieval
10. What methods do you use to search for examples (browse magazines, search the Web, ask friends/colleagues, etc.)?
11. What are the strengths and weaknesses of the methods (listed in question 10)?

Example Storage
12. Do you store design examples (either electronic or hard copy)?
    If yes,
    a. Why do you store examples?
    b. What type of examples do you store?
    c. How do you store examples (bookmarks, file folders, notebooks)?
    d. How often do you review stored examples (can you describe an actual occurrence)?

Example Sharing
13. For what purposes do you share stored examples with others, if at all? Can you describe an occasion when you shared a stored example?
14. How do you share examples with others (email, hard copy, post to the web, etc.)?
15. What types of examples do you share (pictures, sketches, etc.)?

Contribution of Examples
16. Do you ever make your personal design examples (finished products, sketches, etc.) available to the public (online website, blog, etc.)?
    a. What types of examples do you share?
    b. Do you ever purposely not share personal design examples? If so, why?
    c. Is there anything that would encourage you to share more examples?

Figure 15: Interview questions used in Experiment 2
4.2 Results

In this section, I present the benefits and use of examples in the design process structured according to the generic creative process model (A. Warr & O’Neill, 2005), see Figure 1. Next, the results are broken down into retrieval, storage and dissemination techniques. Finally, we discuss designers’ motivation for contributing their own personal examples to the design community.

4.2.1 Benefits of Examples in Design Practice

Examples have many benefits in design, particularly in the preparation and idea-generation phases. Examples were mostly cited in the interviews as being an aid in not only providing a scope of what is already available on the market, but also for providing inspiration for new design ideas. The types of examples and utility of these examples are highly similar. Benefits of examples in design include: they provide a visual framework, allow for reinterpretation of ideas, and are used as a validation tool in the late stages of design.

Preparation Phase- Visual framework

During the problem preparation phase, designers view relevant information associated with the problem in order to develop an understanding of what is required and to provide a basis for generating valid solutions. One of the benefits provided by examples during this phase is their ability to improve communication between the clients and the design team. Clients are often ambiguous and imprecise about exactly what they want. The initial meetings with the client are imperative as they set the stage for the entire project by developing common terms. During
these meetings, clients often provide a report of the competition, projects they like, and directions they are interested in pursuing.

The use of visual examples is crucial for effective client-designer communication because examples allow designers to internalize client needs. As one designer stated, “Design examples are like frameworks. That’s how we should view them. It’s a visual that gets the whole team around an idea or a direction.” Similarly, another designer said, “We use examples all the time. It’s always the first part of the design process, what we refer to as survey and analysis.” Once designers and clients have built shared understanding of terminology and directions, they begin generating ideas.

**Idea Generation Phrase- Reinterpretation of examples**

In the Idea Generation phase, designers must develop novel ideas. Examples have many benefits during this phase such as understanding the current market, reinterpreting designs, and determining the originality of a design. Designers often survey the market for similar products to assess what exists (see Figure 16). For example, one designer explained, “We use examples for most of our projects. Generally we ground our projects with an audit at the beginning where we look at current designs out there to get an idea of what the competition looks like.” This allows designers to ensure their product is unique and to explain why it is unique.

Scoping what is already available on the market can also be beneficial if the design direction has been pursued before: “Sometimes it’s interesting because you find someone has thought of your idea before but they have not developed it. It can be interesting to learn why it wasn’t developed or why it didn’t work.” This information allows designers to move forward in the design process.
or helps guide them in a new direction. It is essential to not only determine if the design idea has already been developed, but also to identify mistakes made in or limitations of previous designs.

The designers reported that examples are also important because they allow for reinterpretation of ideas: “I've always believed in art that everything has been done. Essentially what we are doing is re-appropriating. Why struggle with recreating the wheel when you can just use what you've learned from the process of making the wheel to make it better.” The importance of examples is not limited to merely seeing what has already been created, but also establishing new connections. One designer compared the design process to cooking, “You may not like a recipe, but you like some of the ingredients in the recipe. So you take what you like, maybe add in some new ingredients and create a new recipe.” The same idea applies to design. One Web designer discussed how examples played a role in a website they created; “I see things, it's usually pieces and parts. For example, I saw how Apple shows shadows of their products on their web pages. I took that idea and reflected images on my site to improve the design.” Examples allow designers to identify these connections.

Figure 16: Competitive audit and inspirational pieces from the early phases of the design process.
Evaluation Phase

In the Idea Evaluation phase, the designer evaluates the novel ideas produced and judges their appropriateness and creativeness. “Without examples you are just running blindly. It gives you a starting point but it also gives you rationale for what you've done and why it's a valid solution.” Another designer commented, “Essentially that's what I think I am, a problem solver.” For those who view design as problem solving, it is important to reflect on previous examples as one method to ensure they have created an original and valid design solution. Designers also use examples as a reference in the late stages of design to see where they started and how their design evolved.

Potential Limitations of Examples

Some designers may be concerned that viewing examples contaminates their minds and restricts their creative outputs. In some ways, seeing examples could reduce the range of possible design solutions because it guides the mind in certain directions. This notion of design fixation has been found in empirical studies (A. T. Purcell & Gero, 1992), (N. Bonnardel, 2000). Although this may be an issue, there are disproportionately more designers who feel the benefits of examples outweigh the costs. One designer mentioned they are very influenced by what they see, do, and hear, and wonder if they should divorce themselves from examples. However, they said they “thirst for knowledge” and think it would be hard to shut off and not pay attention to examples. “I think that (looking at examples) sort of walks the line between being an artist and being a designer. Designers are able to pull connections between things and see connections and in art you don’t have to do that. You can kind of follow your own path. We need to know what is going on.”
One way the design firms keep examples from hindering their designers is to bring in designers not associated with the current project to the brainstorming sessions. It’s good to bring those people in because they don’t have the same reference points, haven’t looked at the examples, and therefore bring new perspectives to the project. This keeps the design team from being “so entrenched in the process that they can’t get outside the box.” Other firms handle this differently. When looking for examples they specifically do not look in the same field they are working in. For example if they are doing a packaging design for an office supplier, they won’t look at the products too similar or the “best in the market” because it could hinder their creative output.

4.2.2 Example Retrieval Strategies

There are two ways designers find relevant examples; through active search, when the designer is looking for a particular object or has a specific information need; or through passive search, when the designer is looking for inspiration. Designers use magazines, books, the Web and physical product libraries to find relevant design examples.

Examples in the same design domain were cited as being most useful, but designers reported utilizing many other types of examples. For instance, designers often research the target audience to see what types of culturally relevant artifacts they might be interested in (music, media, and other things users might have around) and try to identify relevant examples. They also look at examples of what’s available in the market (research existing designs; get an idea of how they work). This means if they are working on a packaging project they aren’t necessarily looking only at other examples of packaging, but also areas such as fashion design and current style trends. The following sub-sections will further explore how examples are collected and utilized for different projects in product, graphic and web design.
Product design

The most useful types of examples in product design are examples showing physical form and function. In the preparation phase, product designers collect many samples of current products or take digital photos of products on the market. Because it is nearly impossible to collect all samples of competing products, designers also choose to draw inspiration from other visual sources such as online blogs and stock photo sites.

One project described during the interviews was the design of furniture using environmentally friendly materials. The designer referenced different physical material examples to understand how they could be shaped and found examples of both furniture and products through image searches and Web browsing. They also drew inspiration from a poster series by Lester Beall, an artist famous for his print media. For example, the designer said, “I looked to him to say I’m not going to copy what he did, but I’m going to use his sort of style to then mimic that so people make the connection.” The designer used the concept behind Beall’s designs - simple geometric shapes and a simple color palette – as well as similar products on the market to draw inspiration.

As can be inferred from this case, product designers utilize visual sources to help develop the form, function, and color palette utilized in the final product. Given they are 3D designers, physical examples are always the best, but it’s “almost impossible to get your hands on all of them.” Due to this restraint other sources such as magazines, books, and digital images can be used to draw inspiration and aid in the connection of ideas for future designs.
**Graphic design**

Graphic designers utilize a wide range of examples due to the variable nature of their work. The designers interviewed in this category worked on projects ranging from consumer packaging to annual reports. Both physical as well as other non-tangible sources were utilized by these designers. This domain differs from product design because the designers are not only interested in visual sources. For example, when designing a visual logo, designers need not only images, but background information and rationale for those images because the logo represents the entire company.

One designer described a recent project where they were creating a logo to represent a Christian organization. They wanted the logo to represent the ideals of growth and began browsing gardening sites and researching the history of different types of natural trees. They then examined how different tree limbs branched down and intertwined with each other and how the roots were grounded. They used this as an analogy to represent the idea of giving back to the church and the connection of the church community. This draws on an important difference between graphic design and the other domains; graphic designers must thoroughly understand and be able to explain what their symbols mean because the symbols represent the identity of the company.

![Figure 17: Library of physical components and products.](image-url)
Another facet of graphic design is consumer packaging. Graphic designers that work on packaging projects tend to use physical examples as well as other visual inspirations. Some graphic design firms specializing in packaging maintain a physical library of packaging examples, see Figure 17. This library contains many different types of packaging examples from a variety of sources (deodorant, cleaning supplies, alcohol, etc.). At the start of the design process, designers utilize physical examples as well as inspiration from other visual sources (magazine clippings, books, and Web images). Packaging design is similar to product design in that designers utilize more physical examples because the end product is tangible.

Another type of work tackled by graphic designers is print media. One designer discussed their use of examples in the design of an annual report, “I always look into how I can change the annual report: I always have parameters that are the same every year such as they want a lot of picture, but then they give me an un-godly amount of text. So I look at how I can make lots of text and lots of pictures look good. I like the treatment of this example (see Figure 18) because there is a lot of text there, but they break it up in a way that doesn’t seem too gaudy.” This shows that designers use examples not only for visual inspiration, but also for the design of new and concrete layouts.

Figure 18: Example used in the design of a report.
Web design

Web design often involves aspects of graphic design such as when branding a company and developing a logo as part of a site. Therefore, when the designers are involved in the graphic design aspects of a website, examples are utilized and retrieved in the methods already discussed. However, when they focus primarily on website design (structure, layout, interaction, etc.), the examples collected are quite different. One designer mentioned that when they started designing websites they tried not to look at any other website designs because it is easy to “fall into the rut of making your design look like something else.” However, since there are so many usability guidelines, they state there are far fewer organic Web designs possible.

One difference between Web design and the other design domains is there is a standard set of ingredients. As one designer explained “Almost all websites are the same. They have a header that tells you who they are, a navigation bar either at the top or down the side, a story to get your attention, a lot of ‘news stuff’ and then fillers and links to other pages.” The designers in our study stated they weren’t “recreating” web design, instead they look at other websites, choose the pieces they like and use them to inspire new designs. For example, one designer stated, “I flip through sites that sell (website) templates. I never buy them, but I look at the content. When I look at websites I think, ‘I like this, I don’t like all of it, but I like some of it’. So I pick what I want (layout, colors, fonts) and use it to inspire my design. The design process for me is about figuring out what’s going into the soup (navigational bars, links, news information) and then making soup.”

In addition to the layout of the website they also often look at current or emerging trends in the domain for which they are designing. For example, one designer described the re-design process of a collegiate sports website. They reported visiting other university sports sites, flipping
through related magazines (ESPN magazine), and browsing through trendy digital images on the Web site: Deviant Art. The Web designers stated they rarely initiated an active search for inspiration or to aid with design, but more often found inspiration by passively browsing different visual sources.

**Comparison of retrieval strategies**

Designers utilize many sources for retrieving examples, including the Web, magazines, books, and products from local stores or via a physical product library. Designers have different perspectives about the most beneficial retrieval technique. One designer with seventeen years of experience stated they rarely use the Web to search for examples: “It’s a generational thing. I'm online all the time, but I don't like to be distracted. Going to websites really distracts me. But sitting down and grabbing a book, that doesn’t. It's more visual. It's more physical.” In contrast, another designer said they love the Web because “it's free and constant knowledge”. He added, “I’m usually looking for obscure ideas. The internet is good for that.” Although the views of the best retrieval strategy may vary among the design community, several benefits and limitations became apparent throughout the course of the interviews.

When performing an active search on the Web, designers often struggle to articulate the keywords. One designer said, “It’s weird the way Google searches stuff. Sometimes it will be so easy, you'll type in ‘football logo’ and you'll have things at least you can decipher. Sometimes football logo isn’t the best example because when you say football in the US it’s different than football to the rest of the world.”

Another designer cited they were often too literal when searching the web, “If I’m working on sunglasses, I look at too much sunglasses and not enough ‘sun’ or related words. It’s very useful
for me when I start looking at real abstract terms.” One strategy designer’s use to aid in the selection of keywords is to write down a stream of consciousness for the design topic. For instance, if they are designing a line of baby products they may think of the word ‘purity’ which may lead them to ‘water’ which may lead them to look up different water bottle packaging since they are often centered around the ‘purity of their water’.

The Web also does not provide designers with an idea of what is coming next. As one person said, “You can do an initial skim of a magazine and get a sense of what is happening next. If you click on an example on the internet, you’re taken on a completely different tangent. Books and magazines are more organized.” On the contrary, other designers dislike physical books and magazines because the information is typically filtered by editors and the public view of what is “acceptable for publication.”

Many designers argue the ‘randomness’ of the Web allows for more connections and more inspiration than traditional retrieval strategies; “I like it (the Web) because I start meandering. When you get to page 20 (of a Web search), some weird stuff comes up and it’s awesome because it’s only kind of related to the project or term you searched for.” Most designers feel magazines and books are too restrictive, whereas the Web allows users to easily switch between active and passive search.

Many designers referenced problems ranging from the linear nature of web searches to the lack of a hierarchical structure allowing them to focus more attention on relevant designs. They also have problems with losing information. For example, users might perform a web search, but didn’t book mark it or save it and then can’t find it again. One designer commented on this phenomenon, “I find when you get engrossed in what you are doing you forget to have the bookmarks and the tracking so you can actually find your way back to the point in which you got
distracted by 'the bright shiny objects' and forgot about what you were actually supposed to be doing and follow that up."

A benefit of the Web is the abundance of constantly updated information. They mentioned that they have to search many magazines to get the spectrum of information available to them from the Web. Designers also like that they can search by content such as news or images easier than in books or magazines. They also frequent web blogs, particularly when they are related to design. They visit these sites on a daily basis to gain inspiration and get up-to-date information on current trends in design. Since there are numerous design blogs, designers often utilize RSS readers, or a Web feed to publish frequently updated works.

4.2.3 Example Storage Strategies

There are many types of examples stored by designers including; magazine clippings, copies from books, digital images, physical products and websites. Since designers store almost as many examples as they retrieve, underlying organization is needed in order to quickly retrieve and share them. One designer discussed this problem, “I don't have a problem with the volume of information or finding examples, but there is a problem with how you access it, store it and record it; how you get to the kind of things that would be useful at the time that you need them. There certainly is a difficulty in how to manage information especially now that there is so much around. Dealing with the volume of information has become quite a large task.” Each designer / firm has their own way of categorizing information. There is also a difference in the organization and storage of tangible vs. intangible (electronic) examples.
Tangible Examples

There were two types of tangible examples, actual products and other paper sources (magazine clippings, print-outs, photocopies). Many designers flag relevant examples in books and magazines and put the book/magazines in a pile or on a bookshelf. There are many problems with this strategy; designers forget that they flagged examples, forget why they flagged them, or forget to review them. One designer commented on these problems, “I might go back a couple weeks after I flag an example and say ‘why did I flag it or what about it did I find particularly worth flagging?’ and when I figure that out, then it's something I can build off of when I start the next piece of a design.”

The value of these examples resides only within the designer’s memory of what they stored and why they stored it. One way to combat this problem is to classify tangible objects into categories: “Collecting images and magazines is something a lot of designers do. It's always a problem trying to remember what you've got or to find things or classify [examples] without it taking an extraordinary amount of time classifying them.” One designer posts a reminder on a wall when they collect an example to help them remember what examples they have stored.

Figure 19: Storage of printed electronic examples.
Some designers are more extreme and categorize every tangible example they have. One designer prints all electronic sources and catalogs them into a physical library, see Figure 19. Every couple of months they organize their examples. They have many different file folders such as, “branding (examples or reports on branding), how to treat case studies, corporate ids, colors, diagrams and conference materials.” This strategy is more useful than stacking examples because the designer can easily locate relevant design examples and share them with other designers. However, there is a major shortcoming in the time involved categorizing and filing the images.

**Intangible Examples**

Many digital examples are also collected during the process including: digital images, web blog text, and web pages. There are many structures for organizing these types of examples into a searchable database. Some design firms take pictures of all of the physical samples and categorize them based off of their features in an online system. For example, if they have a shampoo bottle, they save it in the directory ‘hair care’ and then the sub-directory ‘shampoo’. They then name the file with as many descriptors as possible, “We usually start off with the brand and go from there. It’s basically a brain dump.” The advantage of this method is it makes the library searchable by keywords and allows designers to view the library from any computer in the office. The designer added, “If we are careful about the file name, we can find the images with an easy search. For example, if I type in purple or shampoo to a search, I’ll find any products matching that description in our database.”

This is also the main drawback of this type of system – the file naming structure. In one particular firm, there are only two designers who add images to the database in order to keep
consistency in the naming conventions of the file. One designer mentioned, “It's downfall is if I type out ‘3 quarter’ versus ‘3Q’ which is our abbreviation for a three quarter view, it's not going to come up if someone types in one or the other. So that's one of the reasons we've kept it to only two people naming the files to keep it consistent.”

Another designer commented on the use of image libraries, “We don't generally have a library of images, because the general feeling is they aren't very searchable. You are better off just starting over on every project and just dragging together inspiration.” In addition, the database cannot be constantly updated because it is limited by the availability of the people maintaining it: “if we get busy, the images pile up and become unmanageable.”

This type of categorization is often utilized by individual designers. One designer explained, “We rely on the categorization and folders to find where the images are stored on the computer, the name of the image don’t matter. We always keep the same file name as found on the internet.” Designers do not want to take the time to rename a file, but merely place it in a folder relevant to the design. For instance, if it is an image of a coffee mug they may place it in the ‘dish’ category and the sub-category of ‘mug’. Although this classification system is seemingly well organized and reduces the load on the designer, there are some drawbacks. “We use a lot of stock photos and save them in a project folder, but the image is only saved as a generic file name. When we are working on a new project we often say, ‘you know what would be great, photo x’, but you can't find it. It's not efficient; you have it (photo x) but you have to go back and manually search each folder.”

Another problem with this method is designers rarely revisit these folders. As with these tangible files, designers often forget what examples they have stored. In fact, they mentioned only revisiting directories if they are adding a new image to it because it reminds them about the
types of images they have stored. They then browse these folders using the ‘preview’ function, but they still have to search through a large quantity of images.

Other types of examples are saved as links to Web pages, but designers mentioned several problems with re-visiting them. The first problem is distrust in the links, “the Web is always changing so I feel like even if I bookmark it, the link could expire.” To bypass this problem, designers pull images or related information off the Web and save it on their computer. Another method for saving websites is to archive the page. This allows designers to revisit pages if the web address has changed or disappeared. However, the most common strategy is the use of bookmarking websites.

Most designers mentioned they have an abundance of bookmarked sites. “The electronic bookmarks are much like what I do with the magazine bookmarks (flags).” In fact, some designers have the exact same difficulties they have with storing tangible examples; they can’t remember what they have or why they bookmarked a particular site. The ease of bookmarking sites often creates many difficulties, “I do a lot more bookmarking than I refer back to. If there’s something I like I’ll put it in a bookmark.” In order to organize these bookmarks, designers often use social bookmarking sites such as del.icio.us as it supports search and tagging which helps them track their bookmarks. Designers like these sites because they are accessible from multiple locations and allow storage and categorization.

However, there are some problems with these sites. One problem designers have with sites such as del.icio.us is there is no structure which separates the more important bookmarks; all sites are given the same weight. Designers stated they would like to have some mechanism that allows them to indicate the more salient bookmarks. Additionally, since there is a large quantity of sites stored, few are re-visited. As one designer stated, “It (del.icio.us) was annoying because I
had to log into it. I like the idea I can access my bookmarks anywhere, but it was just annoying me. I also didn't like having to tag it. I just want to drag it, use it and run. I know the tags are nice for finding them later, but sometimes you only have a list of five things so it's not worth it to tag it. All of my tags are the same, 'design blog.' The underlying problem with storing examples is the lack of tools enabling designers to be able to quickly categorize and organize examples at their own preference so they are easy to retrieve and review later.

4.2.4  **Example Sharing**

Designers reported that they frequently share examples with colleagues by sending e-mail messages with a link to the example and a brief description. One problem with this technique is that the messages may fill up or ‘spam’ a colleague’s inbox. Another problem is that current systems do not weight the importance of the messages. One way to combat this problem is to create some type of file sharing repository, but this too has its limitations such as people do not want to contribute or do not visit the repository.

One designer described a physical wall that people pinned examples to. The problem was that no one ever looked at it. The designer noted, “I don't want to contribute (examples) if no one looks at them.” Designers prefer email over other sources because they know people are more likely to review and use the example. They like the feedback they get from other designers when they find something particularly useful or inspiring, “maybe if there was a news feed like on Facebook where it gives you statistics of how many people have viewed your example, how many people liked it, maybe I would prefer another method.”
4.2.5 Contributing Personal Examples

We asked about designers’ contribution of examples to web blogs and other design outlets as well as their motivation for doing so. There are several restrictions on the types of examples designers can contribute to these sites, largely due to the confidentiality of their designs. It often takes companies several years to develop a product after its initial design. This means designers are held to confidentiality agreements about the products until they are launched and are not allowed to share any examples related to the work. Designers stated they would like to share their work with others if it was not a matter of intellectual property (IP).

There are also personal reasons for not wanting to share examples. For instance, designers may be concerned that someone will copy their work without permission while others are self-conscious due to the fear of criticism (Ardichvili, Page, & Wentling, 2003). There is also a general consensus that sharing examples will affect their ‘bottom line.’ One designer commented, “There is interesting discussion around the design community about ‘open-source’ because in some ways designers earn their money for the service of thinking of ideas and developing ideas. People find it hard to get their head around the idea that you would ‘give things away’ or share them without limitation.” One designer who contributes examples discussed their rationale, “In the beginning I wanted to put stuff out there. I liked that I didn't know who would see it. It’s also a way to broadcast a bit. I contribute examples for the same reason I add pictures to Flickr; I like to share it with other people around the world.”

Although there are mixed feelings about contributing examples to the public, many designers feel they would be more willing to do so if the process was more simplistic. “I know I should do it (contribute examples), and I need to do it. I get more interested in doing it the simpler the tools are to use.” Designers also mentioned they would contribute more examples if they were aware of people interested or working on similar projects. If they knew their examples were
useful to others, they would be more open to sharing design examples as long as they weren’t violating IP rules.

4.3 **Summary and Discussion**

Example use is a common but enigmatic practice in design. The goal of this study was to understand example use in professional design practice in order to drive the development of design tools that better support these processes. This study outlined the utility of examples in three design domains (Web, graphic and product design), shed light on the benefits and different roles examples serve in different phases of the design process and described how designers search for, store and share examples from physical and electronic sources. The goal of this study was to take the lessons learned from the interview data and translate them into implications for tools to better support the creative process. Therefore, the remainder of this section will outline several implications for how computer-based design tools can better support the use of examples during the design process.

4.3.1 **Augment search to prevent fixation**

Search engines typically assume that users can articulate their information need and have an end target in mind. When designers are actively searching for examples, they can articulate their need, but often do not have a target in mind (i.e. there is no “that’s it”). Not only did designers cite the information at the end of the search results to be most inspiring, but studies have shown that viewing only similar examples increases fixation (A. T. Purcell & Gero, 1992). Therefore, the linear or cluster-based display of search results based on similarity (e.g. (Zeng, He, Chen, Ma, & Ma, 2004)) may not always be optimal for example search.
One way to overcome this obstacle is to not always order search results based solely on similarity to the query. For instance, more results could be displayed on the same page allowing designers to see more examples without having to navigate to pages deemed ‘less similar.’ A second method is to mix dissimilar with similar results to create a more diverse set. To find examples that are less similar but still useful or inspirational, systems could extract results that are lower in ranking for the query term and mix them with the front of the results or include results that are related to the topic of the query at different levels of abstraction. For example, if the designer searches for ‘sunglasses,’ terms relating to ‘sun,’ ‘beach,’ or ‘eye protection’ could also be included in the query (and results). These tactics would give designers more immediate access to less similar but still useful or inspirational results and help alleviate fixation.

4.3.2 Improve capture and visualization of search results

Designers may forget to save interesting examples they find or struggle to re-find those examples later in the process. This is consistent with a recent study showing that forty percent of search queries are attempts to re-find information (Teevan, Adar, Jones, & Potts, 2007). The use of social bookmarking sites can help, but the overhead of using such sites and the dynamic nature of the Web limit their value for designers.

One implication is to improve support for capturing search histories. Though research has begun to address this issue (Amershi & Morris, 2008), when finding examples, designers need solutions that allow them to extract and follow new connections between related concepts and at different levels of abstraction. One solution is to employ techniques for visually structuring the search history beyond the use of linear or hierarchical lists. For example, the history could be organized in the form of a 2D graph on an infinite free-form canvas (e.g. the edges represent queries and nodes represent results), leveraging and extending visual structures often seen in
mind mapping. This would allow designers to build explicit connections among different queries during the example search process, reflect on the state of the search to identify new paths, and pursue new tangents without losing previous search context.

4.3.3 **Integrate physical and digital sources of examples**

Our study shows that designers use electronic and physical sources of examples but have different storage strategies for each modality. Some designers try to merge the two modalities by either taking pictures of physical examples or printing electronic sources. The implication is designers need better synergy between physical and digital sources of inspirational materials. For example, a designer could take digital photos or scan pages of physical magazines or paper sketches and insert and position them in the visual structure previously mentioned. This would not only allow designers to have a unified and holistic example space but also allow them to better retrace their global search process.

4.3.4 **Help designers recall why they stored examples**

Results of our study showed that designers have difficulty remembering why they store examples because they have to store the entire example even when they only found a particular piece interesting. For instance, if a designer likes the visual layout of information on a particular Web page, they are typically only able to store (via a bookmark) the page itself and must remember what they liked about it.

The implication is that designers would benefit from mechanisms that would allow them to quickly and easily record why they found particular examples interesting. For instance, by utilizing a palette of different bookmarking tools, a Web designer could directly mark attributes
of an interesting website such as its layout, font type, or color scheme to aid in the recall of why they stored a link to that page. This basically allows a designer to “tag” the examples with little or no additional effort. Also, if the underlying HTML code is available, design tools could import the components related to the attributes. For example, a tool such as FrontPage could import the layouts of saved Web sites and show them as available templates or could show existing design ideas with the fonts or color palettes previously flagged. This works best for domains where the raw data file associated with an example can be parsed (e.g. the HTML code of a web site). However, similar techniques could be utilized in other domains if designers were able and willing to contribute their raw engineering models.

4.3.5  **Encourage contributions of personal examples**

Designers commonly search for and use examples but rarely share their own personal designs, even though these designs could serve as inspiration for other designers. This imbalance is often due to the fear of real or perceived criticism or lack of motivation (Ardichvili, et al., 2003). The implication is that tools should include mechanisms that help decrease costs of example sharing and/or better link the benefits to the costs.

One way to decrease example sharing costs is to more tightly integrate content creation tools with online personal design blogs. When desired, a designer could blog designs directly from the creation tool without the overhead of a context switch. To reduce fear of criticism, blogged designs could be made available anonymously (e.g. not showing the source location or a designer’s alias with retrieved designs). To better link the benefits that other designers receive to the efforts of contribution, systems could record and show how often others have accessed the designs within a designer’s blog or online portfolio, creating a social incentive.
4.4 Future Directions

The use of examples is a common but enigmatic practice in design but this chapter has made several new contributions to understanding this practice in three design domains. First, our results shed light on the benefits and different roles examples serve in different phases of the design process. Second we described how designers search for, store, and share examples from physical and electronic sources. We also discussed reasons as to why designers often do not contribute personal examples to the design community. Finally, based on the results, we offered new implications for how tools can better support example usage in design.

Although this study outlined the importance of examples in the design process, we only briefly studied the limiting factors of example use; namely designers’ views of examples as a constraining factor in design. Therefore it is unclear how, or to what extent, example usage constrains the ideation process because it is unclear what types of examples are actually utilized in professional practice. The following chapter seeks to fill this gap by detailing a laboratory investigation which highlights the types of examples collected during a design task and their relation to studies on design fixation.
Chapter 5

Experiment 3: Example Retrieval

As the previous two experiments demonstrated, examples are a vital part of the design process. In particular, experiment 1 highlighted the utility of actively and passively searching for inspiration while experiment 2 exposed the utility of examples throughout the design process. However, these experiments did not discuss the types of examples designers collect and their relation to design fixation, as discussed in section 2.2.3.

As we know from our previous discussion on fixation effects, design fixation is connected to designer’s expertise, their familiarity with the examples used, and the similarity of the examples to the design task. However, most research on example use has focused on laboratory investigations where designers are specifically given the examples to be used for inspiration. Little is actually known about the amount of ‘fixating examples’ designers themselves select for a given project or the retrieval processes associated with this behavior. The current experiment aims to provide insights into these processes by surveying, interviewing and monitoring professional and novice designers during a design task. In this chapter, we relate our results to research on design fixation and provide implications for the development of example finding tools, and opportunities for future design research. At the conclusion of this chapter, we also introduce a tool developed from the results of the first three experiments.
5.1 Methodology

Our study was developed to answer two main questions:

- How do designers’ example retrieval strategy and example utilization change during the design process? Does this differ across expertise or design domain?
- Are designers creating cognitively constraining environments by selecting highly similar or familiar examples? Does this differ across expertise or design domain?

One of the challenges of our methodology was deciding how to compare data collected for the design tasks between design domains. One approach would have been to have each of the designers perform three design tasks (industrial, web and graphic) and compare performance based on design domain. However, this would have introduced an unfamiliarity bias with the task domain (graphic designers doing an industrial design task) and it would therefore be difficult to interpret the attributes that affected the behavior (their familiarity with the problem, expertise or design domain). A second approach would have been to choose tasks that are domain independent and equally unfamiliar for all designers. The problem with this approach is the tasks are not representative of designer’s typical work. In contrast, our solution was to choose tasks that were equally familiar to each domain and representative of the designers typical design work. The main advantage is that it allows us to compare and contrast their example finding approaches.

This remainder of this section describes the participants, procedure, metrics used for comparison, the experimental design and the results.

5.1.1 Participants

Participants were recruited for this study by emails and posters advertising the research study specifically targeting local design professionals as well as students and faculty in the design and
engineering college. In all, eighteen professional designers in the fields of graphic, industrial, and web design (6 from each domain) participated in our study which lasted approximately 90 minutes. Of these 18, there were 3 self-classified expert designers with a minimum of 5 years of professional design experience (average 14.0, 14.3, 10.3 for graphic, industrial and web design, respectively) and 3 self-classified novice designers (average 3.2, 1.7, 2.7, for graphic, industrial and web design, respectively) in each domain. For confidentiality purposes, participants will be referred to by their domain area (GD, WD, ID), subject number (1-6) and expertise (E or N) throughout this chapter. For instance, the first graphic designer was a novice therefore this designer will be referred to as GD1N. Subjects were remunerated 30 dollars for participating.

5.1.2 Procedure Overview

The procedure for experiment 3 was as follows. First, participants were asked to read and sign an informed consent document and any questions were answered and participants were given $30 for participation. Next, participants were directed to www.surveygizmo.com and asked to complete a 9 question survey. There were three categories of questions including design background, current example retrieval strategies and current example utilization techniques, see Figure 20 for a complete listing of the questions.
1. What is your professional background (choose the one that best describes you)
   ( ) Web Design     ( ) Graphic Design     ( ) Product/Industrial Design

2. I consider myself an expert in this area (see question 3)
   ( ) Strongly Agree     ( ) Agree     ( ) Neutral     ( ) Disagree     ( ) Strongly Disagree

3. How many years of post-educational design experience have you attained?

4. How often do you use examples during the design process?

5. In which stages of the design process do you use examples?

<table>
<thead>
<tr>
<th>Stage</th>
<th>Never</th>
<th>Almost Never</th>
<th>Sometimes</th>
<th>Almost Always</th>
<th>Always</th>
</tr>
</thead>
<tbody>
<tr>
<td>Early: Problem Formation</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Middle: Idea Generation</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Late: Evaluation and Implementation</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

6. What type of information do you look for when you search for design examples?

<table>
<thead>
<tr>
<th>Information Type</th>
<th>Never</th>
<th>Almost Never</th>
<th>Sometimes</th>
<th>Almost Always</th>
<th>Always</th>
</tr>
</thead>
<tbody>
<tr>
<td>Color Palettes</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Form/Function</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Layouts</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Type Fonts</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Templates</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Other</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

   a. If you selected other, please explain.

7. How often do you use the following to find examples?

<table>
<thead>
<tr>
<th>Resource Type</th>
<th>Never</th>
<th>Almost Never</th>
<th>Sometimes</th>
<th>Almost Always</th>
<th>Always</th>
</tr>
</thead>
<tbody>
<tr>
<td>Magazines</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Books</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Web</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Personal Files</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ask Friend/colleagues</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Other</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

   a. If you selected other, please explain.

8. When using the Web to find design examples I have difficulties:

<table>
<thead>
<tr>
<th>Difficulty</th>
<th>Never</th>
<th>Almost Never</th>
<th>Sometimes</th>
<th>Almost Always</th>
<th>Always</th>
</tr>
</thead>
<tbody>
<tr>
<td>Finding design examples relevant to my design task</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Finding the appropriate key words for a search</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Remembering why I stored a particular design example</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Finding information I searched for before but did not save</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Other</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

   a. If other please explain.

9a. For Web designers: How familiar are you with______?

<table>
<thead>
<tr>
<th>Familiarity Level</th>
<th>Very Unfamiliar</th>
<th>Unfamiliar</th>
<th>Neutral</th>
<th>Familiar</th>
<th>Very Familiar</th>
</tr>
</thead>
<tbody>
<tr>
<td>Designing homepages to increase revenue</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Designing webpages for non-profits</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>National Public Radio</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

9b. For Graphic Designers: How familiar are you with______?

<table>
<thead>
<tr>
<th>Familiarity Level</th>
<th>Very Unfamiliar</th>
<th>Unfamiliar</th>
<th>Neutral</th>
<th>Familiar</th>
<th>Very Familiar</th>
</tr>
</thead>
<tbody>
<tr>
<td>Designing advertising campaigns for campus</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Designing advertisements</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>UNICEF</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Print Work: Posters, brochures, t-shirt designs etc.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

9c. For Product/Graphic designers: How familiar are you with______?

<table>
<thead>
<tr>
<th>Familiarity Level</th>
<th>Very Unfamiliar</th>
<th>Unfamiliar</th>
<th>Neutral</th>
<th>Familiar</th>
<th>Very Familiar</th>
</tr>
</thead>
<tbody>
<tr>
<td>Designing products from recycled materials</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Commonly recycled materials</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Environmentally friendly products</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Figure 20: Survey given at beginning of Experiment 3.
Once the survey was complete, designers were given a task in their domain. The tasks were developed by referencing numerous design blogs and design competition websites. The tasks went through a number of iterations and were reviewed in a pilot study in order to ensure they were representative of typical design work. These tasks were ill-defined in order to create both a challenging problem and allow designers to freely explore the solution space. The tasks were as follows:

- **Develop a new product from recycled material (Industrial Design):** Users were asked to create a new consumer product made from discarded products such as tires, munitions, shipping containers, and cell-phones. They were allowed to choose the focus of their ideas (the products domain) as well as the materials utilized.

- **Create a UNICEF campaign for campus (Graphic Design):** Users were asked to develop an ad campaign for UNICEF in the USA to be implemented on college campuses. The goal of the campaign was to encourage donations to UNICEF. Users were allowed to choose the focus of their ideas (design, type of advertisement, etc.).

- **Re-design NPR.org (Web Design):** Users were asked to develop a new homepage for the National Public Radio website (www.NPR.org). The purpose of the new home page was to encourage donations without making it the focal point of the page. Users were able to choose the focus of their ideas (what the home page should include or how to organize the page, etc.)

Participants were instructed that the goal of the task was not to generate polished solutions, but to collect examples (from any web source they desired) and to generate initial solutions. In other words they were asked to brainstorm ideas that could lead to new solutions and display the ideas in the forms of lists, sketches, notes, keywords, and other formats. They were also asked to store any useful or interesting design examples in a Google Notebook which was demonstrated to the participant prior to the start of the task, see Figure 21 for an image of the Google Notebook application. Participants were then given 90 minutes to collect examples and generate potential solutions.
After the participant had completed the design task, or the 90 minutes had expired, a semi-structured interview was conducted in order to gain a better understanding of designer’s example retrieval strategies. The interview session lasted between 30 minutes and 1 hour. The three categories of questions included: design concepts and idea formation, example finding behavior and solution and task representativeness, see Figure 22 for a complete listing of the questions.

Finally, in order to gauge participant’s familiarity (or fixation) with the examples they retrieved during the study, a post-study familiarity survey was completed on surveygizmo.com one week after experimentation. This survey asked participants to rate their familiarity (0-100 sliding scale) with each of the examples they stored in their Google Notebook on four levels: the website in which the example was retrieved, the creator of the example, the information contained within the example and the design example overall. Before rating the examples, the designers were required to review all of their stored examples in order to minimize the biases of example order. The researchers provided the information for each example (website, creator, etc.) so the participant was only required to rate each example according to the information provided. Therefore, each example received 4 familiarity ratings.
## Design concepts and idea formation

Can you briefly describe your design process for this task from when you were given the task until when you stopped the study?

1. Briefly describe your favorite design concepts generated for this task.
2. Which examples helped you develop this design concept, if any? (Did you save the examples in your Google notebook, why or why not?)
3. Was there any type of example that was particularly beneficial for this task (color palettes, layouts, similar products, etc.)? Why?
4. What types of example were the least helpful, if at all?
5. For what purposes did you utilize examples (background information, re-appropriation, etc.)
6. How and why did these examples affect your design activity?

### Example Finding Behavior

7. Can you briefly describe your strategy for finding examples (start at x.com than do a keyword search etc.)? If so, please explain.
   a. What were the strengths and weaknesses of these methods?
8. Were there any websites that were particularly beneficial for this task? Why?
   a. What type of information did you collect from these websites (pictures, type fonts, templates, etc.)
9. What were some of the difficulties you encountered when searching for examples (keywords, websites to go to, etc.)
10. What other sources of examples or other information would have been useful for this task?

### Solution and Representativeness

11. What was the most difficult aspect of this task, please elaborate (try and discuss many difficulties)?
12. How do you think your process of finding examples could have been improved, if at all?
13. How was this task representative of your typical design work? How was it different?
   a. How familiar were you with this design problem (1-5 5 being very familiar)
14. How was this task representative of your typical process of finding examples? How was it different?

### Google Notebooks

15. How did you decide what information to store in the Google notebook?
16. Were there any features of Google Notebooks that were particularly beneficial for finding design examples?

---

*Figure 22: Semi-Structured interview questions from Experiment 3.*
5.1.3 *Task Representativeness*

The design tasks used in this study as well as the ideation environment were developed in order to create an ‘ideal’ work setting. Idea generation and example finding are sporadic behaviors. Designers will sit down for long periods of time and search for design examples and interleave this behavior with other activities. This study aimed to recreate a typical ideation/example finding session.

According to the participants as well as the patterns of behavior recorded in this study, the designers followed the same patterns of behaviors in this study as they do in situ. GD2N stated, "It [my design process] was pretty much the same because we design in the same process for any project, no matter how big or how small." WD2E mimicked this statement, “It’s pretty similar to the entire design process. You start off with this one notion, 'I like this. I’ll stick here for a while.' So it's pretty similar.” The main difference in the design process was that this example finding behavior is normally distributed across a longer period of time. However, designers stated they felt they generally followed the same overall design process regardless of the time limitations.

Designers were also asked how their example retrieval strategies were similar to or different from their typical process. IDE1 represented the general consensus by saying, “I would say it’s pretty typical. I do rely on the web for the investigation phase of the research. Because in design, a lot of time it’s about finding the right question, not the right answer. To do that, you have to do a lot of looking.”

Therefore, although this study was conducted in a 90 minute design session in a laboratory setting, the design tasks used and the design process followed by the designers were highly representative of their typical processes.
5.1.4 \textit{Metrics}

We computed the following metrics for each subject to measure the similarity of the examples to the design task:

- \textit{Number of unique keywords} is the number of unique unrepeated keywords used during the design task for each designer. For example, if a designer searched for HIV kids, then HIV children, then HIV facts for kids, a total of 4 keywords would be counted (HIV, kids, children, and facts). However, words that were direct synonyms were only counted once. In order to account for this, keywords were confirmed by checking for synonyms using Webster's Dictionary ("Webster's New Millennium Dictionary of English," 2006). In this case, kids and children are direct synonyms. Therefore, the total number of unique keywords in this example is 3 (HIV, kids/children, and facts).

- \textit{Number of keywords from task description}. Once the unique keywords were identified, they were compared to the task description and reused keywords were identified. For instance, if an industrial designer searched for "tires", this would count as a keyword from the task description since the word ‘tires’ is in the description.

- \textit{Percentage of the keywords from task description} divided by the number of unique keywords. This was computed in order to normalize the results.

These computed metrics as well as the results from the post-survey were used in order to understand the amount of examples similar to the design task and familiar to the designer as both measures have been linked to cognitively constrained environments.

The percentage of keywords from the task description was used in order to explore the amount of similar examples designers retrieved. This measure was used because it provides a direct comparison between the task description and the users’ search queries. The higher the percentage of terms the designers used from the task description (i.e. UNICEF, AIDS, HIV, etc.), the more similar the search results will be to the design task.

In order to explore example familiarity, the 4 measures of participant-reported familiarity were used. These metrics allow for a deeper understanding of the variety of both familiar and unfamiliar examples designers use when they are allowed to freely explore the solution space.
5.1.5  *Experimental Design*

In order to answer our first research questions the ordinal data from the survey was coded such that a response of ‘always’ corresponded to a 5, and ‘never’ corresponded to a 1 with the rest of the responses coded similarly. A Kruskal-Wallace test (non-parametric data) was used to compare across domain and expertise (i.e. do graphic designers utilize examples more in the early phases of design as compared to industrial and web designers).

In order to address the second research question, a (M)ANOVA was performed with the dependent variables obtained from (1) the post-study questionnaire for familiarity (website, creator, information and overall familiarity with each of the examples) and (2) the calculated similarity measure of percentage of keywords from task description. In both cases the Independent Variables were expertise (novice, expert) and design domain (graphic, industrial and web).

All of the statistical tests were performed using SPSS v 18.0. The level of significance for all statistical tests was 0.05.

5.2  *Results*

Figure 23 shows some of the design concepts created by our participants during the study as well as the examples they collected to inspire these designs. In this section we will compare and contrast the strategies used to retrieve these examples as well as discuss the types and variety of examples designers’ retrieved.
5.2.1 **Example use in the Design Process**

Our first research objective was to understand how the different design domains retrieve examples and how (and when) they utilize them during the design process. Overall, both our qualitative and quantitative results showed differences in the utility of examples during the design process, see Figure 24.

Figure 23: Design concepts created by graphic designers (large image) and examples collected to inspire their design (smaller images). From left to right: GD3E, GD1N and GD4E.

Figure 24: Industrial designers almost always use examples while graphic designers mostly use them in the early phases of design.
For instance, graphic and industrial designers utilize examples more frequently than Web designers in the early phases of design \( (p=0.03) \). In addition, industrial designers are the only design group that utilize examples frequently in every phase of the design process \( (p=0.01) \). From this figure we can also see that graphic designers’ utility for examples decreases throughout the design process; meaning they almost always use examples in the early phases of design and rarely use examples in the late phases. Finally, web designers only occasionally use examples in each of the phases of design. The main focus of this study was to understand how examples were retrieved and used in the early phases of design.

In order to gain better understanding of when examples were retrieved by our participants and how this was related to their ideation sessions, we logged the start time and duration of example retrieval as well as idea generation sessions for each of the designers in our study, see Figure 25 for a sample log. This figure shows those graphic designers’ often retrieved examples when starting the task but rarely return to the examples after they started ideating. On the other hand, industrial designers often began by searching the web to find background information and then
interleaved ideation and example finding sessions. Finally, the web designers tended to start ideating solutions immediately and only looked to the web to find examples when they had specific solutions in mind. The following sections will outline more specifically how and when examples are retrieved and used in each of the design disciplines from a qualitative perspective.

**Graphic Design**

Graphic designers collected between 10 and 37 examples (mean 19) during this study. The designers typically began their design process by understanding the UNICEF organization by searching the web for background information and examples of previous UNICEF campaigns. As GD2N stated they weren’t looking at previous UNICEF campaigns to see what they should do, but rather to avoid replicating what had already been done. In other words, they were not using the existing UNICEF designs as inspiration, but as a means to gain a better understanding of the project space. Other designers specifically avoided other campaigns that UNICEF had done before or any related genres mainly to avoid clichés. GD5E stated, “if someone is too familiar with your design (if it’s too similar to what’s out there now), then it’s going to be ignored. So, I always try and figure out what’s new or what’s different, just to get someone’s attention.” In other words, they specifically avoided similar designs and instead started by searching for background information on UNICEF.

After finding background information the designers relied on image search to find additional examples. The keywords they used to perform the image searches were often ones they wrote down while reviewing background information on UNICEF (i.e. love, care and donations). Many of the designers also used an online thesaurus to help identify additional keywords. GD5E provided rationale for this behavior, “I think the way designers work is that one word connects to another word which connects to another word. For me, the thesaurus serves that purpose. If I
find a word that is a 'hit', or a new direction, I'll use it in a search... I also look for antonyms. Often the opposite of what I'm looking for gives a more visual concept and is the most direct."

The designers then used these keywords identified to search for relevant examples. GD2N stated, “Mainly I was doing word searches that related to what I was hoping to do. Those word searches, most of the time, got me somewhere useful.” Specifically, GD2N mentioned looking at the Meth project because they felt this campaign did a good job with their ‘branding’, or people’s ability to look at the poster and know what it represented. S/He wanted to create the same kind of atmosphere in this new UNICEF campaign.

GD3E, on the other hand, drew inspiration from the Red Campaign posters because it is a campaign still being used. Like GD5E they were looking for not only what they wanted to do, but also what they wanted to stay away from. They stated, "The red campaign affected (my ideas) a lot because I wanted to go the exact opposite route of that. I do not want to be remotely connected to it", (see Figure 23c). This means some designers looked for images that were the opposite of what they are looking for to draw inspiration.

After the designers performed their example searches, they began brainstorming possible design solutions. They drew inspiration from past projects and examples they had retrieved. For instance, GD4E came up with the idea of doing a poster campaign about uneducated women; “Somehow in my search I typed in 'veiled women', I don't really know how I got to that point, but I got this image and it was powerful and she looked like she was a wreck and I came up with the idea that 'uneducated women come out as hurting women”, reusing the images s/he found on Google, see Figure 23a. On the other hand, GD1N drew from their background in university athletic poster design. After s/he looked at the current UNICEF posters, GD1N felt they could
use some of the ideas for the university, focusing on using basketball and football players in the poster because s/he felt most students look up to those players.

After the designers started ideating solutions they only returned to the web when they had a specific target in mind. GD2N stated their rationale for this behavior, "I usually am looking for specific images for a project after the brainstorm... I know what I’m looking for." GD1N also discussed their process, "I came up with the idea to do basketball players holding children’s hands, but I thought that might be a bit cheesy. While I was searching for examples I came across this image (see Figure 23b) which has nothing to do with my idea but I said 'OK, the UNICEF logo is round,' let’s combine them." At this point they went back to the Web to find a UNICEF logo to fit into their concept. This also shows that although the designer was searching for a particular artifact, they were also open to serendipitous discovery during the process.

In sum, graphic designers start the design process by searching for background information and examples of current campaigns. They then search for images of concepts they had developed and ideate potential solutions with the retrieved examples. They mainly look for examples in order to gain background information, find inspiration, and find examples for re-appropriation.

**Web Design**

Web designers collected between 4 and 19 (mean 9) design examples during the study. Figure 26 illustrates some of the examples collected from WD3N to aid in his/her design. The web designers typically started the design process by visiting the current National Public Radio (NPR) website and identifying potential problems. Next, the designers immediately started ideating by replicating the existing design either on paper or using Adobe Photoshop and applying their knowledge to solve some of the problems they identified.
They then began looking for existing design solutions for the unsolved problems. WD6E commented on this strategy, “My first question is, 'what is wrong with the current site'. I start looking at what the problems are and then I brainstorm about what the known solutions are. One of my design philosophies is that there are designers out there that have already knocked the problem out. So, if you can build off of an existing standard, you’ll save a lot more time” Therefore, in order to get an idea of how other websites combated the identified problems, the designers started searching for similar examples.

In this part of the design process, web designers are focused primarily on the layout of the page and therefore were looking primarily for examples of different layout structures. WD3N said the web pages they use (their design examples) depend on their familiarity with the particular design field, in this instance non-profit news websites. S/he said, “If it's something I'm aware of I'll typically use the websites that I think are most common or most frequently used.” WD4N commented that one limitation to only using these familiar or frequently used websites is that
they are not as open to new perspectives. However, WD4N felt they had enough familiarity with NPR that they didn’t need to search for more examples. The most frequently utilized sites for this task were CNN, PBS, and Yahoo. They used these sites to compare their features and organizational strategies to that of NPR. The rationale for choosing these websites was that they were “winning designs that had stood the test of time.” WD5E also went to sites like CSSVault and CSSbeauty to look for other designs that people thought were ‘pretty’ or ‘nice’, noting what they like about each design and applying it to their concepts.

Only one designer, WD1E, started looking for these ‘known solutions’ by searching for similar websites using terms such as non-profit sites, best non-profits in the world, and what makes a good non-profit site. WD1E said, “I wanted to know what it is about a non-profit site that makes it successful. I didn’t know what it was about non-profit sites because I have never done one before.” However, because the search results ‘weren’t what they expected’ they went back to the sites they normally use for inspiration (in this case CNN and Yahoo).

Once the designers find the example web pages, they often open them in browser tabs, and refer back to them when necessary during their design process. They only close these examples when they feel they are no longer useful.

In sum, web designers start by applying their own design knowledge to the design task and then either search for examples or utilize the websites they deem ‘best in the field’ as design inspiration. Finally, they only return to the examples they have already retrieved, and have kept open in tabs, once they have started ideating to compare different design features. The examples they retrieve in the early phases of design tend to deal primarily with the layout of the site and things they can reuse in the new design.
Industrial Design

Industrial designers collected between 12 and 20 (mean 17) examples during the study. They typically began the task by searching for relevant background information on recycled products to draw inspiration. ID5E said, “I went to LOHAS.com to get information on market shares. They gave me a good idea of the areas that have bigger markets. That is where I got the idea of working on product packaging”. Similarly, ID4N first searched for images and recycling statistics on tires and then, “tried to use the existing knowledge to relate it to things that the tires could be used for.” Another starting strategy designers had was to connect the design task to things they had recently seen. For instance, ID1E started by retrieving an article s/he a colleague earlier in the day about RV recycling. ID4N recalled a recent television show that used black piping to heat water so they started researching how they could use this in their design. They used this memory to give them a starting point for their research.

Once the designers had a firm understanding of the problem space, they began visiting familiar websites and weblogs to aid in ideation. ID6E stated, “Forums are good once you have knowledge because they give you more context. If you go to something that is only text based or only image based, it’s hard to find something insightful.” ID2N actually started searching Google before visiting the blogs without much avail. “I started with Google images and typed in water bottle but I got back 400 results. I'm not going to sift through that. Plus I started seeing stuff that wasn't really relevant so I started to think of things I had done in the past and design blogs I had visited, so I went there.” After designers reviewed the designs on these websites and ideated a potential solution, they then performed searches examples similar to their solutions.
For instance, ID3N started his/her design process at instructables.com and treehugger.com because s/he was familiar with those websites. While browsing the website s/he found many examples of products reusing plastic bottles which led him/her to look into glass bottles because s/he thought it was a more versatile material. His/Her first idea was to create a cap kit or pour spout for different bottle sizes so that people could reuse bottles by filling them with soap and using them as a soap dispenser. S/he then looked for examples similar to this design concept to help her understand if, and to what extent, this idea exists. While looking at these examples s/he thought, "This would look nice in a garden because they're natural colors." So s/he started looking for examples of bottle fencing, something s/he could design and customers could buy in a store where there would be a little stake in the ground and the top of the bottle would be partially buried, see Figure 27.

ID1E had a similar strategy. S/he first started by trying to search for recycled RV’s to get more background information on what has already been done. However, s/he accidentally mistyped it
and searched for ‘recycles RV’. Although this wasn’t what s/he was looking for s/he called it a ‘happy mistake’ because it took him/her in a new design direction; creating a line of RV accessories made from recycled RV parts. At this point s/he went to a familiar website, campingworld.com, and searched through the accessories on the website. S/he felt it was important to see the visuals because, "A word is going to take me down a narrower hallway where a picture gives me a hallway where if I push against it, it might open up another door." This shows that although the background information in the forms of articles is definitely important, industrial designers also need visuals to help them ideate more solutions.

In general, Industrial Designers tend to search for relevant background information, exhaust internal ideation resources and then intersperse ideation and example finding sessions. During the first ideation session they tend to identify relevant keywords for future research paths. They utilize examples retrieved in the early phase of design to gain background information on the problem and find examples that can be re-appropriated in a new design. In addition, they use examples to help them make new connections between concepts. For instance, ID03N stated “examples help me think. If I see something it might remind me of something I already know. It helps me make new connections.”

5.2.2 Variety of Examples Utilized

Our second research objective was to understand the differences in the variety of examples utilized during the design task. Previous research has shown that designers can become fixated on designs based on their expertise, familiarity with the examples, and the similarity of the example to the design task. Our study aimed to quantify these results by dissecting designers’ search queries and self-reported design example familiarity. In sum, we found that graphic web
designers view more familiar examples than industrial designers, see Figure 28, and expert designers view more familiar and similar examples than novices, see Figure 29.

As a reminder, we used the keywords measures from the task description to compare how similar the design examples were to the design task. Participants also reported their familiarity with each of their stored examples on four dimensions. WD6E was excluded from familiarity calculations due to technical difficulties.

**Creative design domains**

Our first goal was to understand if there were any differences in the variety of examples used between the three design domains. The results show that industrial and graphic designers were less familiar with the creator (photographers, designers, organizations) of their retrieved examples (p=0.01). Industrial designers were also less familiar with the information contained within the examples (i.e. Meth Advertisement Campaigns, news websites, recycled products blogs, p=0.00) and the examples overall (p=0.00). This means that graphic and web designers are utilizing examples they are more familiar with, see Figure 28.
This could possibly be attributed to industrial designers utilizing a lot of design blogs which post design examples from a multitude of sources. Therefore the industrial designer would be familiar with the website, but since the information is constantly updated, they would be unfamiliar with the information contained within the site. There were no significant factors for similarity.

**Expertise**

Our second goal was to understand if there was a difference in the variety of examples used between expert and novice designers. Our results (see Figure 29) show that expert designers were more familiar with the examples they collected (based on overall familiarity p=0.02), and the information contained within the examples (p=0.00). In addition, experts utilized a higher percentage of keywords from the task description, meaning they searched for examples similar to the task (p=0.03).

![Figure 29: Average familiarity/ Similarity by expertise. Novice designers view less familiar examples.](image-url)
These differences could be attributed to experts being more successful than novices when searching for design examples, specifically because they are more able to articulate their search terms (White, Dumais, & Teevan, 2009). Additionally, many of the expert designers related the task to previous projects in which they had topic familiarity which has also been linked to an increase in search performance (Kelly & Cool, 2002). Thus, the chance of expert designers encountering less familiar or less similar search results could be greatly reduced due to their search patterns and the highly filtered and ordered search results.

5.3 Summary and Discussion

Using examples to inspire creativity is a highly valued and prevalent method in the design industry. However, it remains unclear how designers search for and utilize examples to stimulate creative outputs and how technology could best facilitate this behavior. Our work has made several contributions addressing this problem by performing a laboratory study with representative and challenging design tasks in three domains.

First, we observed notable differences of search strategies and example utilization across design domains. The variety of resources designers seek and dimensions of examples they use are highly dependent on the design domains. Second, we found that the design domain and expertise influence the examples designers utilize during the design process. This phenomenon is partially due to the design fixation effect. We found that expert designers rely on their own knowledge and only seek highly familiar or similar examples, losing the potential benefits from exploiting a more diverse set of examples. Graphic and web designers also collected more ‘fixating’ examples than industrial designers. This leads to interesting research questions on how we can better support these design domains that have received relatively little attention in terms of design fixation in research literature.
The results obtained in this study allow us to derive practical design implications for tools aiming at facilitating example finding activities.

5.3.1  *No one size fits all solution*

One of the most important implications derived from this study is that there is no ‘one size fits all’ solution to aiding designers in example finding. Our results show differences in the type and variety of examples designers choose as well as their strategies for retrieving these examples across design domains and levels of expertise. Therefore, the type of constraining environment (too many similar or familiar examples) varies based on these dimensions. In addition, the purpose for retrieving these examples varies during the design process. For instance, industrial designers use examples in the early phases of design to gain background information whereas web designers are looking for known solutions to existing problems. In turn, it is difficult to construct a global interface that retrieves a suitable set of examples for all designers because the same set of examples does not have the same degree of appropriateness for all design situations.

Although there is no one solution, one possible aid for example finding could be to create a visual display that shows how constraining a set of examples are. For instance, this could be accomplished by creating a window in the web browser that shows the last few examples collected, the similarity of those examples to one another and the familiarity of the examples to the designer. These measures could be calculated using the provenance data of the examples and the browsing history of the designer. This type of system would benefit designers by creating a cognitive awareness of fixation effects.
5.3.2 *Allow for serendipitous discovery through example re-composition*

One of the main uses of examples is to take existing designs and compare the features between the products in order to identify interesting components (N Bonnardel, 1999). However, our results showed that designers in different domains seem to have various strategies for comparing examples. For instance, web designers open similar websites in multiple tabs, in order to tab between the websites and compare functions (see what works and what doesn’t).

One of the main problems with current methods for comparing and analyzing features of existing designs is that they examples are often stored in a linear structure or designers are required to flip between Web pages. Neither of these strategies truly allows designers to make interesting comparisons between examples. If better comparative techniques were developed, designers would be more open to serendipitous discovery during this comparative process.

For instance, GD1N had this type of experience when they were searching for examples of basketball players holding children’s hands. They said, “I came across this image (Figure 23b) which has nothing to do with it (my idea) but I said ‘OK, the UNICEF logo is round,’ let's combine them.” Recognizing the similarities in size and shape of these objects allowed GD1N to make the association between the basketball shape and the UNICEF logo shape.

One way to aid in these discoveries is to show examples at different level of abstractions. For instance we could apply segmentation techniques for extracting basic shapes and objects from the stored examples and make them available as elements in existing content creation tools, e.g., as layers in Adobe Photoshop. Designers could then arrange the layers to quickly synthesize new design ideas or form new connections between the elements.
5.3.3 Allow for ‘happy mistakes’

Stumbling upon unanticipated information has been shown to be a valuable resource for designers. In fact, all of the designers in this study quoted having ‘happy mistakes’ or discovering examples by chance. One designer even compared this process to gambling stating that if you keep ‘throwing your dice’ (searching) you’ll eventually win (find a good example). However this statement also outlines one of the big limitations of using the web for example finding: either the web is too random and provides too many irrelevant results, or it is too specific not allowing for serendipitous discovery.

In the beginning of the design process, novice designers become overwhelmed by the amount of results appearing from search engines and therefore tend to collect a large variety of design examples which do not fit a particular theme. ID04N stated they “have a problem when there are too many things to look at... it’s easy to get distracted.” GD6N said, "internet searches waste time and are distracting because I see something really cool and it leads me to another blog or link instead of focusing on the task.” On the other hand, expert designers primarily utilize search when they are looking for a specific object utilizing only familiar examples. Both of these cases can lead to constraining environments as the expert designers are relying on familiar examples and novice designers aren’t able to focus their design ideas.

One way to combat these problems is to create a search plug-in that allows designers to rate their confidence in their search terms or what they are looking for. As their confidence decreases, the more dissimilar results appear. The more confident they are, the more specific the search results will be. This would allow designers to have an example set that better matches the expectations or needs of the design situation.
It should be noted that websites and interfaces like StumbleUpon which deliver web results that match the user’s personal preferences are not necessarily ideal for designers because their design goals and preferences are constantly changing. For instance, IDN2 began looking at shipping crates, then glass bottles, and then garden fences.

5.3.4 Design Education and Design Research

One implication of our findings for design education is promoting the use of social networks in the early phases of design. Although our results show that the web is the most frequently used source for example finding, industrial designers also stated that they often refer to colleagues and friends to aid in example retrieval. Something that may be beneficial for designers, especially novice designers, is to view the impact of examples collected by peers in the early phases of design. Therefore, future studies should research the usefulness of examples retrieved from peers in the early stages design and their impact on the creative process. Our results cited many instances where novice designers had a difficult time finding a starting place for their research. The use of social networks could potentially aid in this process.

5.4 Summary Part I

The main goal of Part I of this dissertation was to develop guidelines for tools that can better support the design process through an in-depth understanding of existing example-based design methods. Through these studies we were able to determine various methods for ideation and identify examples as the cornerstone of the design process. Our results have shown that designers like using the Web to find examples because the ‘randomness’ and resourcefulness of the Web allows them to make more connections and draw more inspiration than traditional sources like books and magazines. However, there are two main issues designers identified with
using the Web for inspirational searching in our studies: it is either too random providing too many irrelevant sources or too specific not allowing for the discovery of inspirational content which can lead to design fixation.

One of the causes of designers search difficulties is the sheer amount of information available on the Web. Search engines return hundreds, if not thousands of Webpages per query. With the ever increasing amount of information available, it is becoming more and more difficult for designers to parse through the information to locate design inspiration and develop new search directions. To compound this issue, the goal of most search engines is to return highly specific and ordered results making designers privy to the artifacts directly related to their search term reducing their chance to find inspirational content. In addition, after many search sessions, designers may become fixated on a particular topic and struggle to develop new search and design directions. For these reasons, much of the design implications in Part I of this dissertation proposed ways we can alter existing technologies or develop new Web-based technologies that aid in example finding. In particular, these implications were identified for altering Web-based example finding methods:

*Experiment 1: Supporting Active Search, Supporting Passive Search*
*Experiment 2: Augment Search to Prevent Fixation, Improve capture and visualization of search results*
*Experiment 3: Allow for “Happy Mistakes”*

As a capstone of Part I of this dissertation we realized some of implications listed above to improve inspirational searching on the Web, see Figure 30. We developed TweetSpiration, a Web-based design tool that displays filtered tweet results and socially derived word associations based on a user-entered term. This application builds off of the ideas outlined in Chapter 2 in that it develops new cues to trigger relevant memories and aid in ideation. By using socially-
derived word associations instead of semantically related word associations we can help in the triggering of more distantly related images in memory, and thus help in the search for ideas in associative memory process. In other words, by filtering Twitter content, designers can gain a new perspective of the design problem which can lead to new search directions, directions they may not have thought of without using the tool. TweetSpiration can be used at any time, but it is particularly beneficial when designers have difficulty developing new search terms or are looking for new search directions.

Figure 30: TweetSpiration, a Web-based tool developed to aid designers in inspirational searching on the Web.

Although not discussed in this dissertation, an initial study of this application revealed that users think TweetSpiration helps them develop new inspirational search directions on the Web and provides new perspectives on the design problem. In fact, six of the eight pilot users said
they would ‘definitely use the application’ again for inspirational searches, see Herring et al. (2011) for in-depth discussion of the tool and study design.

In sum, Part I of this dissertation provided an in-depth understanding of existing design practices and highlighted ways we can extend existing design tools, or develop new design tools to aid in the design process. Of particular interest in Part I of this dissertation was the use of examples during the design process. Part II extends the interest developed in Part I on example usage, but instead of focusing on example-use in ideation practices, we now focus on how examples can be used in an educational setting to help identify ones design skill levels.
Part II: Design Skill Assessment
Chapter 6

Literature Review: Design Skill Assessment

Part I of this dissertation focused on understanding example-based ideation practices and identifying ways in which we can aid designers through the development of computer-based design tools. Part II of this dissertation still focuses on the design process and the three phases of analysis, generation and evaluation, but instead of looking at how we can aid individual in this process, it focuses on how we can objectively measure ones design skill level in each of these phases. Since researchers have noted the importance of evaluating design skill not just by one’s ability to develop quality solutions, but also by their ability to *successfully* perform each step of the design process (McKenzie, 2002), it’s important that we access individuals in all of these areas.

The second part of this dissertation focuses on design skill assessment from an engineering design education perspective. The current chapter is meant as an outline of current assessment techniques in engineering design education. The chapter will begin with a discussion of typical engineering design course structures as well as an overview of different types of learning objectives. Following this discussion I will outline assessment techniques currently utilized in engineering design education. The next chapter will discuss a new approach to design skill
assessment and chapters 8-10 will detail three studies developed to test this new objective approach in an engineering design course.

6.1 **Engineering Design Course Structure and Learning Objectives**

In engineering design courses, students gain knowledge of design concepts and principles through lectures and tutorials on a variety of subjects important to the design process (Eggermont, Brennan, & Freiheit, 2010). These courses are typically organized such that the first half of the course is focused around teaching design principles and the second half of the course is focused on teaching design methodologies and the application of the learned principles in either labs, semester projects or some combination thereof. In these courses, open-ended design projects (or labs) serve as a catalyst for learning as these projects require students to harness and *apply* the knowledge he or she has gathered during the semester (Davis, Trevisan, Beyerlein, Harrison, & Thompson, 2007).

When we look at the objectives of these engineering design courses we find that the types of skills being taught require drastically different knowledge levels, with some tasks requiring only memorization to complete and others requiring sophisticated analytical skills and creativity. For instance, a human factors engineering design course at the University of Illinois states the course objectives as:

1. Recognize good and bad designs and understand their contributing factors;
2. Use appropriate human-centered techniques and methodologies to analyze human needs, design technological solutions, and evaluate designs; and
3. Identify common design problems and suggest solutions as part of a usability evaluation cycle (Richard, 2010).
These objectives exemplify both memorizing design principles as well as developing solutions to complex problems which require different skill levels for completion.

In order to classify drastically different learning objectives according to their required skill levels Benjamin Bloom and colleagues formulated a system referred to as Bloom’s Taxonomy of Educational Objectives (Bloom, 1956). This taxonomy divides educational objectives into three domains: Cognitive, Affective and Psychomotor. The cognitive domain revolve around knowledge, comprehension and critical thinking while the affective domain focuses on the way people react emotionally and the psychomotor domain describe one’s ability to physically manipulate a tool or instrument. Although a goal of Bloom’s Taxonomy is to have educators focus on all three domains, traditional education tends to focus on the Cognitive Domain and skill set. The learning objectives in the Cognitive domain are (6 being the highest level):

1. Knowledge: Can the student recall or remember the information?
2. Comprehension: Can the student explain ideas or concepts?
3. Application: Can the student apply learned information to solve a problem?
4. Analysis: Can the student break things down into their elements and formulate explanations for findings?
5. Synthesis: Can the student create something or combine elements in novel ways?
6. Evaluation: Can the student make and justify value judgments or selections from among alternatives?

The 6 levels in the taxonomy are hierarchical such that each level is subsumed by the higher levels. Put another way, a student functioning at the ‘analysis’ level has also mastered the material at the ‘application, ‘comprehension’ and ‘knowledge’ level.

Since its creation, over 20 updates or revisions of Bloom’s Taxonomy have been developed. Perhaps the most closely associated with Bloom’s original work is that of Anderson et al. (2001), a former student of Bloom’s. According to Anderson, his revision of Bloom’s original taxonomy was needed to update the framework in terms of 21st century cognitive psychology research.
while using a more common language and articulating more realistic examples. Unlike the original Taxonomy, this new taxonomy involves two basic dimensions instead of three including the knowledge domain and the cognitive process domain. The knowledge domain involves four types of knowledge (factual, conceptual, procedural, and metacognition) while the cognitive process domain involves six types of thinking, see Figure 31.

Comparing the cognitive process domain with Bloom’s original Taxonomy, we see this revision uses verbs instead of nouns to describe the different levels of the taxonomy and also inverts the first and second objectives in the hierarchy (evaluation and synthesis, respectively), see Table 4 for explanation of levels. In addition, while Bloom’s original taxonomy was structured hierarchically, the revised taxonomy only considers the three lowest levels as hierarchically ordered, while the top three levels are parallel. These top three levels (analyze, evaluate and create) are known as higher-level thinking skills and most engineering course instructors will state that these are the skills they desire their students to master during their course (Felder & Brent, 2004). However, in order to evaluate student mastery of these levels, instructors must select appropriate assessments.

Figure 31: Revised Blooms Taxonomy which uses nouns instead of verbs and does not consider the top three levels of the taxonomy to be hierarchical.
<table>
<thead>
<tr>
<th>CATEGORY</th>
<th>COGNITIVE PROCESS</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Remember</strong></td>
<td>Retrieve relevant knowledge from long-term memory</td>
</tr>
<tr>
<td></td>
<td>RECOGNIZING (identifying)</td>
</tr>
<tr>
<td></td>
<td>RECALLING (retrieving)</td>
</tr>
<tr>
<td><strong>Understand</strong></td>
<td>Construct meaning from instructional messages including oral, written and graphic communication</td>
</tr>
<tr>
<td></td>
<td>INTERPRETING (clarifying, paraphrasing, presenting, translating)</td>
</tr>
<tr>
<td></td>
<td>EXEMPLIFYING (illustrating, instantiating)</td>
</tr>
<tr>
<td></td>
<td>CLASSIFYING (categorizing, subsuming)</td>
</tr>
<tr>
<td></td>
<td>SUMMARIZING (abstracting, generalizing)</td>
</tr>
<tr>
<td></td>
<td>INFERRING (concluding, extrapolating, interpolating, predicting)</td>
</tr>
<tr>
<td></td>
<td>COMPARING (contrasting, mapping, matching)</td>
</tr>
<tr>
<td></td>
<td>EXPLAINING (constructing models)</td>
</tr>
<tr>
<td><strong>Apply</strong></td>
<td>Carry out or use a procedure in a given situation</td>
</tr>
<tr>
<td></td>
<td>EXECUTING (carrying out)</td>
</tr>
<tr>
<td></td>
<td>IMPLEMENTING (using)</td>
</tr>
<tr>
<td><strong>Analyze</strong></td>
<td>Break material into its constituent parts and determine how the parts relate to one another and to the overall structure or purpose</td>
</tr>
<tr>
<td></td>
<td>DIFFERENCIATING (discriminating, distinguishing, focusing, selecting)</td>
</tr>
<tr>
<td></td>
<td>ORGANIZING (finding coherence, integrating, outlining, parsing, structuring)</td>
</tr>
<tr>
<td></td>
<td>ATTRIBUTING (deconstructing)</td>
</tr>
<tr>
<td><strong>Evaluate</strong></td>
<td>Make judgments based on criteria and standards</td>
</tr>
<tr>
<td></td>
<td>CHECKING (coordinating, detecting, monitoring, testing)</td>
</tr>
<tr>
<td></td>
<td>CRITIQUING (judging)</td>
</tr>
<tr>
<td><strong>Create</strong></td>
<td>Put elements together to form a coherent or functional whole; reorganize elements into a new pattern or structure</td>
</tr>
<tr>
<td></td>
<td>GENERATING (hypothesizing)</td>
</tr>
<tr>
<td></td>
<td>PLANNING (designing)</td>
</tr>
<tr>
<td></td>
<td>PRODUCING (constructing)</td>
</tr>
</tbody>
</table>

Table 4: The Revised Blooms Taxonomy which includes verbs instead of nouns and places 'create' (synthesis) as the highest learning objective instead of evaluate (evaluating).
6.2 **Assessment in Engineering Design**

Assessment of student achievement in design is an important part of engineering education. However, engineering programs face special challenges in assessing students design capabilities and providing meaningful feedback because engineering design is mostly subjective in that there are no mathematical proofs or conclusive experiments to grade (Davis, Gentili, Trevisan, & Calkins, 2002). In light of this difficulty, educators have developed a variety of ways to assess student performance during design courses typically involving some combination of written reports, presentations, quizzes, prototypes, peer evaluations or evaluator judgment (Sobek & Jiain, 2004). The following sections will focus on outlining these assessments, the skill levels they assess according to Bloom’s Taxonomy and the utility of these assessments in engineering design education.

### 6.2.1 *Objective Assessments*

One might believe that in order to test ones design skill we might merely provide a quantitative test of design knowledge because the more knowledge a student gains, the more design skill they will develop. For instance, asking student’s close-ended questions (i.e. questions with right and wrong answers) on design principles or the design process could serve as a means to quantify student design knowledge. These types of assessments have a long list of positive benefits: they are easy to score, they avoid reliability with problem raters, and they can be assessed at the individual level (Bailey, 2006). Carefully written questions can even tap into the process-related design knowledge of students. However, this approach only taps into students ability to memorize design principles, parts of the design process and design concepts and thus students who score well on these assessments do not necessarily have the ability to apply this knowledge (J. R. Anderson, 2005). In fact, Oxman (2004) once stated that it is naïve to assume that the
more knowledge a student gains, the more design skill they will acquire because we often observe the contrary; the more knowledge a student gains, the less design skill they acquire.

This lack of correlation between objective measures of design knowledge and design skill can be attributed to these types of assessments focusing primarily on the lower levels of Bloom’s Taxonomy. In other words, these assessments do not measure student’s ability to analyze evaluate and create solutions – all of which are an important part of the design process. Since what is of interested to engineering design educators and future employers is not only the rehashing of design principles, but the ability of students to reach the higher-tiers of Bloom’s Taxonomy, different assessment techniques are required to evaluate students in these areas. This is not to say the memorization of design principles is not an important aspect of engineering design education, it just brings to light that these assessments on their own are not sufficient means of assessing a student’s ability to reach the higher levels of learning.

In order to overcome some of the deficits of using close-ended questions, some researchers have started utilizing open-ended questions on their assessments (Sims-Knight, Upchurch, Pendergrass, Meressi, & Fortier, 2003). These questions can include asking students to reflect on a design task via a survey or essay or they may be asked to map ‘all the words they associate with the concept design’ (Newstetter & Kahn, 1997). In both of these cases, students are given the freedom to construct responses that show knowledge at levels beyond pure memorization of design processes and concepts. In other words, the open-ended nature of these problems allows students to construct responses that show knowledge from multiple levels of Bloom’s Taxonomy. However, both of these cases (essays, and associated words) have one major challenge that must be addressed: how do we objectively assess responses? The problems in objectively grading these types of open-ended question relates in large part to the second type of assessment techniques used in engineering design: subjective assessments. The following
section will discuss how instructors deal with the open-ended nature of these types of responses as well as other types of subjective assessments.

6.3 Subjective Assessment Measures in Engineering Design

In addition to tests of design knowledge, instructors often use design projects and labs as both a catalyst for learning as well as a means to evaluate student design skill. During these projects and labs, students are required to analyze both the design problem as well as current solutions to the problem, develop ideas based on their analysis, and evaluate and select the best solutions. Although these projects are critical for teaching and evaluating design skill, the open-ended nature of these projects makes it difficult to develop objective criterion to assess performance. Because of these difficulties, many design project and lab grades are structured around evaluations performed by the course instructor, a design ‘expert’ (Demirbas & Demirkan, 2007). One of the main issues with this assessment method is that design tasks do not typically have concrete right or wrong answers the end result and therefore must be subjectively graded by the course instructor. These judges, however, are subject to cognitive biases and limitations based on their own beliefs and expectations.

Subjective bias is defined as decision making or evaluation based on personal, poorly measurable, and unverifiable data or feelings weighted against objective, unbiased data which impede the ability of individuals to make valid decisions or evaluations (Dyro, 2004). In particular, evaluators can be biased based on the order in which they evaluate the projects (cue primacy and anchoring effects), or their inability to look at the problem from a perspective other than their own (framing bias) (Wickens, Lee, Liu, & Gordon Becker, 2004). In order to overcome some of the biases involved in subjective grading, a number of techniques have been developed such as grading rubrics and peer evaluations (Davis et al., 2009).
In order to improve grader reliability, instructors often utilize scoring rubrics. Rubrics define criteria for assessment, qualities that will be assessed and identify the levels of performance students should demonstrate for each of the quality outcomes before products are created (Asunda & Hill, 2007). These rubrics serve as both a guide of expected learning outcomes for students as well as a tool for instructors to structure their evaluations, see Figure 32 for rubric example. Although these scoring rubrics can help mitigate subjective biases in grading by providing specific criterion for assessment, they are time intensive to develop and they also increase the time to grade project and lab deliverables. In addition, subjective biases can still have a major influence the scoring or rating of each criterion.

Another way to mitigate instructor bias is peer evaluation. Student involvement in assessment has been increasing in recent years, particularly in engineering (Falchikov & Goldfinch, 2000). In these activities, students are engaging in criteria or standards and applying them to make judgments on peer projects. Although instructors are often concerned with the degree of agreement between their marks and those awarded by the students, a meta-analysis of research on peer and teacher remarks found that peer assessments align with professor assessments when students are asked to provide an overall judgment on well understood criteria instead of judgments on several individual dimensions. Although peer assessment can be utilized as a
substitute for instructor feedback student biases can still effect the evaluation of the design outputs.

In summary, while there are a variety of assessments that have been developed for engineering design, most objective assessments are untested for reliability of validity and thus instructors often have little confidence in their assessment of student competence in design (Davis, et al., 2006). In addition, while projects and labs allow for the development of the upper-tiers of learning, subjective evaluations are often utilized to evaluate student performance reducing their effectiveness. Therefore, new assessments are needed to provide an objective evaluation of these upper-tiers of learning.
The Bayesian Truth Serum Assessment

As Chapter 6 brought to light, one of the difficulties in accessing design skill in educational and professional settings is developing objective criterion and as such, many design course grades are structured around evaluations of student design projects performed by the course instructor, a design ‘expert’ (Demirbas & Demirkan, 2007). Similar to the lab and projects given in the educational setting, individuals are often given design tasks in professional settings (i.e. interviews) as a way to access design skill (NPR, 2003). For example, at Microsoft, job applicants are often provided with design tasks such as “Design an alarm clock for deaf people” which is intended as a means to access individual design thinking abilities (Ace The Interview, 2010). The ideas developed by candidates in this process are then subjectively judged by the interviewer. One of the main issues with this assessment method, as well as the ones discussed in the previous chapter, is the evaluator must subjectively measure the performance. These subjective evaluations are problematic however, because there is no public criterion for assessing judgmental truthfulness (if the individual’s judgment of the quality of the design idea is correct). Thus, it brings to question if there are other ways we can access design skills without expert input and when objective truth is unknowable and how, if at all, examples can be utilized in this process.
One way to use examples in the objective evaluation of design skill is to test one’s ability to identify examples that embody principles of good design. The more design skill one has acquired, the better they are at identifying pitfalls in previous designs, or designs that do not incorporate principles of good design (N. Bonnardel, 2000). Therefore, it can be concluded that designers with more design skill will be more able to identify examples that embody principles of good design. If this theory holds true and an individual’s ability to rate design examples is correlated with their design skill, example usage could be extended to not only inspiring design ideas or teaching individuals through example designs, but also objectively testing one’s design skill. In other words, individuals design abilities would no longer be appraised based on expert evaluations, but instead would rely on an individual’s ability to rate examples. The use of examples for this type of evaluation also seems pertinent in design education, as Experiments 1-3 highlighted the importance of selecting quality examples during the design process. One potential way to remove subjective bias and utilize ratings of example quality is through the use of the Bayesian Truth Serum.

7.1 Bayesian Truth Serum

The Bayesian Truth Serum (BTS) is a method for eliciting subjective information when objective truth is intrinsically or practically unknowable (Prelec, 2004). The method consists of an information scoring system that induces truthful answers from a sample of rational (i.e. Bayesian) expected value-maximizing respondents. This method does not assume that the researcher knows the probabilistic relationship between different responses making it applicable to previously unasked questions. In addition, since no understanding of probabilistic relationships is required, the method can be applied by a researcher that is a complete outsider of the domain.
Unlike earlier approaches to ‘test theory without an answer key” (5), the Bayesian Truth Serum does not privilege the consensus answer but instead assigns high scores to answers that are more common than collectively predicted, with predictions drawn from the same population that generates the answers. These responses are deemed ‘surprisingly common’ and the associated numerical index is called an information score. Those who have superior meta-knowledge – knowledge of others opinions – receive higher BTS scores and thus are considered to have more knowledge than others in the sample population. This adjustment in the target criterion which does not privilege the consensus answer removes respondents bias towards answering what they feel is the likely group mean. In other words, truthful responding remains the correct strategy even for someone who is sure their response represents a minority view.

Figure 33: Example used in Study 1 where participants were asked to judge the design examples relation to the salient feedback principle

7.2 BTS Method

To calculate a BTS score, participants are provided with a series of survey questions and asked to provide personal answers and also a prediction of the percentage of the sample population that they feel will endorse each answer. For example, $n$ participants are provided with questions which have $m$ answers: i.e. Which of the following statements best describes Figure 33 (m=3):

(a) This is an effective example of the visual salience design principle
(b) This example is not relevant in reference to the visual salience design principle
(c) This example depicts a violation of the visual salience design principle
Respondents are asked to provide a personal answer (i.e. a-c) as well as a prediction of the proportion of the sample that will endorse each answer: e.g. I think 30% of other respondents will think this it is an effective example, 2% will say it is not relevant, 68% will think it’s a violation. Once the answers and predictions are collected, a BTS score can be calculated.

7.2.1 BTS Calculation

The calculation of the BTS score involves four steps. First, Let $x_k^r \in \{0, 1\}$ indicate whether respondent $r$ has endorsed answer $k$ and $y = (y_1^r, \ldots, y_m^r)$ be her prediction of the sample proportions ($y_k^r \geq 0$, $\sum_k y_k^r = 1$). The BTS algorithm then proceeds in four steps.

**Step 1** - Calculate the average $\bar{x}^k$ of the endorsement and the geometric mean $\bar{y}^k$ of the predictions:

$$\bar{x}^k = \frac{1}{n} \sum_{r=1}^n x_k^r, \quad \log \bar{y}^k = \frac{1}{n} \sum_{r=1}^n y_k^r$$

**Step 2** - Calculate the BTS score of each individual $r$:

$$u^r = \text{information score + prediction score}$$

$$u^r = \sum_{k=1}^m x_k^r \log \frac{\bar{x}^k}{\bar{y}^k} + \sum_{k=1}^m x_k^r \log \frac{y_k^r}{\bar{x}^k}$$

**Step 3** - For each answer $k$, calculate the average BTS score $\bar{u}^k$ of all individuals endorsing answer $k$:

$$\bar{u}^k = \frac{1}{n \bar{x}^k} \sum_{r=1}^n x_k^r \bar{u}^r$$

**Step 4** - Select the $k$ that maximizes $\bar{u}^k$

The prediction score calculated in Step 2 scores the accuracy of respondent’s predictions of the empirical distribution of answers. In other words, it scores how well the respondent’s predictions match the empirical frequencies. The best prediction score is zero, attained only when prediction exactly matches reality, $y_k^r = \bar{x}^k$. The information score, on the other hand, looks at respondents personal answers giving a higher score to those whose answer is ‘surprisingly common’ or more common than collectively predicted. For example, if an answer is endorsed by 30% of the population ($\bar{x}^k$) but the population predicted only 5% ($\bar{y}^k$) of
respondents would select that response then those who selected that answer would receive a high information score because their response is ‘surprisingly common’ \((\bar{x}^k > \bar{y}^k)\). On the other hand, if the predictions averaged 75% \((\bar{y}^k)\) and only 30% of the population \((\bar{x}^k)\) selected the answer, the answer would be deemed surprisingly uncommon and those who selected it would receive a low information score \((\bar{x}^k < \bar{y}^k)\).

7.2.2 **BTS Indexes of Expertise**

BTS can also be used to identify expert respondents if knowledge correlates across multiple questions (Prelec & Seung, 2010). BTS suggests two possible indexes for rating the expertise of a single respondent: the individual and pooled index. The *individual index*, is the BTS score \(u^r\) averaged across all responses while the *pooled index*, is the average BTS score of the answer endorsed by respondent \(r\) averaged across all questions, \(\sum_{k=1}^{m} x_r^k \bar{u}^k\). The pooled index filters out individual differences in prediction competence (i.e. long run calibration of predictions).

7.3 **Overview of Experiments**

The Bayesian Truth Serum has been validated both theoretically and empirically and has proven to be a solid way to identify experts when subjective judgments remains the only source of evidence available, and there is a possibility that most people may be wrong (Prelec, 2004; Prelec & Seung, 2010). Since design is very subjective, the Bayesian Truth Serum could provide a way to accurately predict genuine design knowledge and design skill levels. In order to test the applicability of BTS for this domain, three studies were conducted:

**Experiment 4:** Can BTS properly identify expertise in human factors design?

**Experiment 5:** Can BTS properly identify one’s ability to develop quality design ideas?

**Experiment 6:** Can BTS properly identify one’s ability to provide useful design feedback?
7.4 Summary

The three studies conducted in part II of this dissertation seek to understand how, or to what extent, student design skill can be identified through example-centric assessment practices and a novel application of the Bayesian Truth Serum. More specifically, these studies seek to understand how, or to what extent, this method can be used to assess higher levels of learning (analysis, evaluation and creation) which transpire throughout the three phases of design: analysis, idea generation and evaluation. In order to test BTS’s applicability in this domain three studies were conducted. The following chapters outline these studies in detail.
Chapter 8

Experiment 4: BTS Skill Assessment

While there are a variety of assessments that have been developed for engineering design, most objective assessments are untested for reliability of validity and thus instructors often have little confidence in their assessment of student competence in design (Davis, et al., 2006). In addition, while projects and labs allow for the development of the upper-tiers of learning, subjective evaluations are utilized to evaluate student performance reducing their effectiveness. Therefore, new assessments are needed to provide an objective evaluation of these upper-tiers of learning.

In this chapter, we introduce and describe a study we developed to test the applicability of BTS for assessing design skill levels in engineering education. This methodology is intended to support cases where human judgment is the only source of evaluation available and there is a possibility that most judges may be wrong.

8.1 **Methodology**

The current study was developed to generate a BTS score for each participant based on their ability to evaluate design examples in terms of human factors principles and predict the proportion of the study population that would endorse each answer which can be used to
compare BTS to measures of genuine knowledge and response accuracy. This section describes the participants, the survey developed, metrics used for comparison, the experimental design and the results.

8.1.1  *Participants*

Students were recruited for this study from the Human Factors Engineering Design Course at the University of Illinois. In all, 106 students (3 sophomores, 59 juniors and 44 seniors) with less than 1 year experience in Human Factors Design participated in this study. Students participated on a voluntary basis and the study took approximately one hour to complete. Each student was provided with a unique 3 digit ID to ensure anonymity. In addition, the instructor for the course also participated in the study as our resident ‘expert’.

8.1.2  *BTS Survey Design*

In order to test student’s ability to rate design examples in terms of human factors principles and predict the empirical distribution of responses from the sample population, a 30 question survey was designed. For each question on the survey, participants were presented with an electronic picture of an existing product and asked to rate the example as an effective example of a human factors principle, a violation of the principle, or one in which the principle does not apply, see Figure 34. Hereafter these survey answers will be referred to as ‘effective’, ‘not relevant’, and ‘violation’. In addition, to selecting a personal answer for the survey, participants were asked to predict the proportion of other students in the course that would endorse each response in order to judge their meta-knowledge of others opinions.
The following human factors principles were tested during this survey: color coding, comfort, physical affordances, analogies, population stereotypes, safety, salient feedback, shape coding, stimulus-response compatibility and visual salience. These principles were selected because they are core to the IE 340 curriculum and students are expected to be able to identify problems with existing solutions according to these principles when they complete the course. Three pictorial design examples were selected for each of the ten human factors principles for a total of 30 BTS survey questions. The pictorial examples selected for the study were chosen by the experimenter prior to experimentation to exemplify the human factors principle, display a violation of the human factors principle, or depict a design that was not relevant for the human factors principle presented, see Appendix A for an exhaustive list of the questions.

In addition to the 30 question BTS survey, a 10 question objective True/False quiz on the human factors principles tested in the BTS survey was given. The questions can be found in Appendix B.
8.2 **Procedure Overview**

Upon arriving in the laboratory, participants read and signed an informed consent document. Once completed, the following instructions were provided:

“For the following 30 questions you will be asked to rate different design examples based on human factors principles. We ask that you answer the questions as truthfully as possible, even if you think your opinion represents a minority view (i.e. you think others will answer different than you will). In addition to giving a personal answer, you will also be asked to predict the percentage of other students currently taking IE 340/ PSYCH 358 Human Factors Design Course that would endorse each answer. Remember to answer to the best of your ability!”

Students were not graded on this assignment but were asked to respond as truthfully and accurately as possible. Once participants read the instructions, any questions were answered.

Next, participants were directed to [www.surveygizmo.com](http://www.surveygizmo.com) where the survey was located and given 45 minutes to complete the BTS survey with questions presented in random order for each participant. Following completion of the BTS survey, participants were given an additional 15 minutes to complete the ten question objective True/False quiz on the human factors principles tested in the BTS survey.

8.2.1 **Experimental Design**

There were three main questions to be addressed in the first experiment:

1. Can BTS determine the correct answer of survey questions as implied by theory,
2. Does the pooled or individual BTS index of expertise correlate with genuine knowledge, and
3. Does the BTS information score or prediction score correlate with genuine knowledge
In order to answer question 1, we identified the answer \( k \) that maximized the BTS score \( \bar{u}_k \) for each question on the BTS survey (see Step 4 of BTS computation) and compared to the majority opinion for each question as well as the answer selected by the course instructor.

For questions 2 and 3 the pooled index, individual index, information score, and prediction score were tested for correlation with the following measures using SPSS V.18 and the Pearson two-tailed significance test:

- **index of conventional wisdom** (CW): the number of BTS survey questions for which a respondent votes with the majority opinion for that question
- **response accuracy**: the number of times the respondent selected the same answer as the course instructor
- **quiz grade**: the score (percentage) of correct responses to the 10 question objective true/false quiz given at the end of experiment 4
- **test average**: the average score (percentage) of the midterm and final grade in the Human Factors Design Course
- **course grade**: the final grade (percentage) the student received in the Human Factors Design Course

The IE 340 test average and course grade were used as measures of genuine human factors knowledge whereas the quiz grade was used as a representative of the respondent’s knowledge of the human factors design principles tested in the BTS survey.

In order to address the problem of multiple comparisons in experiment 4 and maintain a familywise error rate of \( \alpha = 0.05 \), the Bonferroni correction was used to test each individual test at a significance level of \( \frac{\alpha}{n} \) where \( \alpha = 0.05 \) and \( n \) is the number of hypotheses in experiment 4 (\( n = 18 \)) (Abdi, 2007). The correction ensures that the significance level for the whole family of tests is, at most, \( \alpha \). Therefore, the significance level for each of the individual tests in Experiment 4 was \( \frac{\alpha}{n} = \frac{0.05}{18} = 0.002777 \).
8.3 Results

This section serves as a summary of the results from Experiment 4.

8.3.1 Can BTS determine the correct answer?

Can the BTS algorithm determine the true (or correct) answers as implied by theory (Prelec & Seung, 2010)? Following steps 1-4 of the BTS calculation we find that the BTS decision on the correct answer for each of the 30 questions differs from the majority decision on 9 questions. When comparing the BTS decision to the IE 340’s instructors answer (the ‘correct’ answer) for these 9 questions we find that for 2 questions it corrects a majority decision to align with that of the instructors response (it corrects a wrong majority decision), while for 7 questions it reverses a majority decision that previously aligned with the course instructor (it reverses a correct majority decision). In other words, unlike other studies which report that BTS can improve on majority decisions when applied to individual questions (Prelec & Seung, 2010), our study finds the contrary. So why is it that our application of BTS showed contradictory results?

In the previous study on the applicability of BTS scores for selecting correct state capitols, the average respondent was slightly above chance endorsing 31 correct answers out of 50 (62% correct responses). Our average respondents performed roughly the same as the previous study performing slightly better than chance by selecting 19.5 correct answers out of 30 (65% correct). In the previous study, the collective judgment was slightly better than the average respondent with the majority decision selecting 36 correct answers out of 50 (72% correct). However, in our study differed the collective judgment greatly outperformed the average respondent’s selecting 28 out of 30 correct answers (93% correct). What this means is that for 28 out of 30 questions on the BTS survey, the majority opinion was correct. For meta-knowledge to be an effective truth diagnostic, information must be unevenly distributed among respondents so that some
pieces are widely shared but other pieces are known only by a minority (Prelec, 2004; Prelec & Seung, 2010). In other words, if the majority opinion can select the correct answer in most cases, the utility of BTS may decrease.

Another differentiation between our study and the previous study is that we offered three answer choices (effective, violation and not relevant) for our participants instead of two. If we look at the 2 questions that BTS corrected a wrong majority decision (question 11 and 22) we see in both cases it changed a majority decision of ‘not relevant’ to the answer of ‘effective’, see Figure 33 for a depiction of one of the two corrected survey questions. For this question, respondent answers were relatively evenly distributed among answers with 34% selecting ‘effective’, 31% selecting ‘not applicable’ and 35% selecting ‘violation’. The correct answer, ‘effective’ had the highest ratio of actual versus predicted endorsement (33% vs. 14%) making this answer the most ‘surprisingly common’ of the three choices. This is an appropriate application of the information score in that the ‘surprisingly common’ answer was in fact the correct answer.

For the 7 questions that the BTS score incorrectly adjusted a correct majority decision (questions 2, 4, 7, 9, 15, 18 and 26), it changed the majority answer from either ‘effective’ or ‘violates’ to ‘not relevant’. For these questions, the population predicted that, on average, only 12% would endorse the ‘not relevant’ option when, on average, 30% selected the option. In all these cases, the ‘not relevant’ answer was the most surprisingly common response resulting in a higher BTS score. This result could be attributed to respondents not being able to accurately predict what proportion of the sample would select the ‘not relevant’ option. Another explanation could be that more students did not know the correct answer and thus selected ‘not relevant’, resulting in a higher percentage of respondents that selected that choice.
Table 5: The average percentage of respondents that endorsed each answer ($\bar{x}$) and the average prediction of the percentage of the population would endorse each answer ($\bar{y}$). The ratio of $\bar{x}/\bar{y}$ is used in Step 2 of the BTS calculation to determine ‘surprisingly common answers’. The higher the ratio, the more ‘surprisingly common’ the response.

Table 5 represents the average number of respondents that endorsed each answer over all 30 survey questions. It also shows the average predictions for each answer. As can be inferred from this table, the ‘not relevant’ option was the least popular answer and the population predicted, on average, that less than 10% of respondents would select the option. Therefore, whenever more than 10% chose the option, the answer became ‘surprisingly common’. In other words, the ‘not relevant’ option had the lowest threshold for becoming ‘surprisingly common’ which attributed to BTS incorrectly selecting this option in 7 cases.

Since the surprisingly common score more often than not incorrectly adjusted a correct majority opinion, it was hypothesized that by altering the BTS algorithm to give less emphasis to the surprisingly common score we could improve BTS’s accuracy. In order to test this theory step 2 of the BTS algorithm was altered as follows:

**Step 2**- Calculate the weighted BTS score of each individual $r$:

$$u^r = c \times \sum_{k=1}^{m} x^r_k \log \frac{\bar{x}_k}{\bar{y}_k} + d \times \sum_{k=1}^{m} x^r_k \log \frac{y^r_k}{\bar{x}_k}$$

where $d = 1-c$

Using the newly formulated BTS, we examined how varying the weight (d) of the information score affects the number of correct answers the BTS algorithm identifies. In other words, how
does providing more (or less) weight to the information score affect BTS’s accuracy in this application?

As can be seen in Figure 35, the less weight we provide to the information score, and thus the more weight we give to the prediction score, the more accurate the BTS algorithm becomes. In fact, when we remove the information score altogether (d=0), the algorithm identifies 27 of 30 correct answers. This means that by reducing the weight of the information score, we can improve BTS’s ability to select the correct answer, although it is still not perfectly accurate (when d=0 the BTS score corrects 2 majority opinions and reverses 3 correct majority opinions). While BTS still does not outperform the index of conventional wisdom (which identified 28 of 30 correct answers), by re-weighting the BTS algorithm we can greatly improve its accuracy from 23 to 27 correct answer selections.

![Figure 35: The number of questions the weighted BTS algorithm identifies with varying levels of d, the weight of the information score in step 2 of the BTS calculation.](image)

8.3.2 Do the BTS indexes of expertise correlate with genuine knowledge?

Prior studies have shown that the BTS expertise indexes can be used to identify expert respondents. More specifically, the pooled index score has been shown to be an excellent
predictor of genuine knowledge correlating significantly better than the measure of conventional wisdom with response accuracy (Prelec & Seung, 2010). In order to determine if the BTS indexes of expertise are good indicators of genuine knowledge in our study, we tested the indexes for correlation with the course grade, exam average and quiz score. As a benchmark, we also tested these measures correlation with conventional wisdom and response accuracy. The results from this study can be seen in Table 6.

<table>
<thead>
<tr>
<th></th>
<th>Individual Index</th>
<th>Pooled Index</th>
<th>Conventional Wisdom</th>
<th>Response Accuracy</th>
<th>Quiz Grade</th>
</tr>
</thead>
<tbody>
<tr>
<td>Course Grade</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Pearson Correlation</td>
<td>.222</td>
<td>.091</td>
<td>-.078</td>
<td>-.094</td>
<td>.117</td>
</tr>
<tr>
<td>Sig. (2-tailed)</td>
<td>.022</td>
<td>.354</td>
<td>.425</td>
<td>.336</td>
<td>.233</td>
</tr>
<tr>
<td>N</td>
<td>106</td>
<td>106</td>
<td>106</td>
<td>106</td>
<td>106</td>
</tr>
<tr>
<td>Test Average</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Pearson Correlation</td>
<td>.223</td>
<td>.110</td>
<td>-.066</td>
<td>-.085</td>
<td>.103</td>
</tr>
<tr>
<td>Sig. (2-tailed)</td>
<td>.022</td>
<td>.261</td>
<td>.499</td>
<td>.385</td>
<td>.297</td>
</tr>
<tr>
<td>N</td>
<td>106</td>
<td>106</td>
<td>106</td>
<td>106</td>
<td>106</td>
</tr>
<tr>
<td>Conventional Wisdom</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Pearson Correlation</td>
<td>.302</td>
<td>.594</td>
<td>1</td>
<td>.936</td>
<td>-.257</td>
</tr>
<tr>
<td>Sig. (2-tailed)</td>
<td>.002</td>
<td>.000</td>
<td></td>
<td>.000</td>
<td>.008</td>
</tr>
<tr>
<td>N</td>
<td>106</td>
<td>106</td>
<td>106</td>
<td>106</td>
<td>106</td>
</tr>
<tr>
<td>Response Accuracy</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Pearson Correlation</td>
<td>.360</td>
<td>.666</td>
<td>.936</td>
<td>1</td>
<td>-.280</td>
</tr>
<tr>
<td>Sig. (2-tailed)</td>
<td>.000</td>
<td>.000</td>
<td>.000</td>
<td>.004</td>
<td></td>
</tr>
<tr>
<td>N</td>
<td>106</td>
<td>106</td>
<td>106</td>
<td>106</td>
<td>106</td>
</tr>
</tbody>
</table>

Table 6: Results from correlation test between the BTS indexes of expertise, measures of human factors knowledge, and traditional measures of CW and response accuracy.

Our results indicate that the measures of genuine human factors knowledge (course grade and test average) did not correlate significantly (p<0.002) with the indexes of expertise or the measures of conventional wisdom and response accuracy. However, both indexes of expertise correlated significantly with conventional wisdom and response accuracy. Figure 36 demonstrates the correlations between response accuracy and conventional wisdom and response accuracy and the BTS indexes of expertise. From this figure we can infer that the pooled index is a better indicator of accuracy than the individual index which confirms prior
findings. However, the measure of conventional wisdom has a stronger correlation with response accuracy than both indexes of expertise. Therefore, although the indexes do correlate with genuine knowledge, they do not outperform the index of conventional wisdom.

![Figure 36: Expertise correlates with both BTS scores and Conventional Wisdom.](image)

We hypothesized that we could improve the correlation between response accuracy and the BTS indexes of expertise by performing the same weighting to step 2 of the BTS calculation that we performed in section 5.2.1. In order to test this theory, we calculated the individual index and pooled index using 5 values of d (0, 0.25, 0.50, 0.75 and 1.0) and tested these calculations for correlation with response accuracy, see Figure 37. Similar to the results found in the previous section, we see that by reducing the weight of the information score in both the individual and pooled index calculations, we increase the correlation between the indexes and response
accuracy. In fact, similar to in the previous section, when d=0, we receive the highest correlation coefficient between the measures.

![BTS Expertise Indexes and Response Accuracy Correlation](image)

Figure 37: The correlation between the BTS indexes of expertise and response accuracy at varying levels of d, the weight of the information score in step 2 of the BTS calculation. By reducing d, we can substantially increase the relationship between the pooled index and response accuracy. All correlations significant at p<0.000 except d=1.00 which is not significantly correlated.

8.3.3 *Do the BTS information or prediction scores correlate with genuine knowledge?*

Since the original BTS algorithm incorrectly adjusted a correct majority opinion on 7 questions, and we found that by reducing the weight of the information score we could increase the performance of the BTS algorithm, we wanted to test the individual elements of the BTS algorithm (the information score and prediction score) for correlation with genuine knowledge. Therefore, we tested correlation between the average information and prediction scores against the variables in 5.2.1, see Table 7.
Our results show that the prediction score correlates significantly with the measures of conventional wisdom and response accuracy while the information score does not. In other words, the ‘surprisingly common’ score has no correlation with these measures. However, when comparing the prediction scores correlation ($r=0.37$ with response accuracy) with the weighted BTS algorithm ($r=0.82$ between response accuracy and the pooled index when $d=0$), we find that the prediction score does not correlate as strongly with these measures. In other words, the prediction score on its own is not as strong of a measure of genuine knowledge as the weighted BTS algorithm.

<table>
<thead>
<tr>
<th></th>
<th>Course Grade</th>
<th>Test Average</th>
<th>Conventional Wisdom</th>
<th>Response Accuracy</th>
<th>Quiz Grade</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Information Score</strong></td>
<td>Pearson Correlation</td>
<td>.155</td>
<td>.111</td>
<td>-.033</td>
<td>.066</td>
</tr>
<tr>
<td></td>
<td>Sig. (2-tailed)</td>
<td>.114</td>
<td>.257</td>
<td>.737</td>
<td>.501</td>
</tr>
<tr>
<td></td>
<td>N</td>
<td>106</td>
<td>106</td>
<td>106</td>
<td>106</td>
</tr>
<tr>
<td><strong>Prediction Score</strong></td>
<td>Pearson Correlation</td>
<td>.209</td>
<td>.217</td>
<td>.324</td>
<td>.370</td>
</tr>
<tr>
<td></td>
<td>Sig. (2-tailed)</td>
<td>.032</td>
<td>.026</td>
<td>.001</td>
<td>.000</td>
</tr>
<tr>
<td></td>
<td>N</td>
<td>106</td>
<td>106</td>
<td>106</td>
<td>106</td>
</tr>
</tbody>
</table>

Table 7: Results from correlation test between the BTS algorithm information and prediction score components, the measures of human factors knowledge, and traditional measures of CW and response accuracy.

### 8.4 Summary and Discussion

The goal of the current experiment was to determine BTS’s ability to select the correct answers on a survey, and its ability to identify expert respondents in an educational setting. First, the discussion will focus on answering the experimental questions posed earlier in the chapter.
8.4.1 *Can BTS determine the correct answer of survey questions?*

The BTS calculation was able to identify 23 of 30 correct answers, but it was not as accurate as conventional wisdom which identified 28 correct answers. Because the majority opinion in our study selected most of the correct answers, we hypothesized that the ‘surprisingly common’ score was actually harming BTS’s ability to identify the correct answer. To test this theory, we re-weighed step 2 of the BTS algorithm and found that by reducing the weight of the information score in BTS’s formulation, we were able to enhance its accuracy. In fact, when we removed all weight from the information score, BTS was able to identify 27 correct answers. In other words the ‘surprisingly common’ factor, or the information score, actually hindered BTS’s ability to correctly predict answers because the majority was correct most of the time.

8.4.2 *Does the pooled or individual BTS index of expertise correlate with genuine knowledge?*

When comparing the BTS indexes of expertise to the measures of genuine human factors knowledge (response accuracy, course grade and test average), we find that neither index correlated significantly with the course grade or test average (but the individual index was trending on significance), but they both correlated with response accuracy. In other words, those with higher BTS indexes had more accurate responses. Although the correlation between the pooled index and response accuracy was noteworthy ($r=0.67$), the traditional measure of conventional wisdom was more strongly associated with response accuracy ($r=0.94$). This is due to the fact that the majority opinion was correct on 28 of 30 questions.

We again hypothesized that the information score, or the surprisingly common score, actually hindered BTS’s ability to predict genuine knowledge and therefore tested varying weights of the information score in the BTS formulation for correlation with response accuracy. Our results
showed that the less weight we give to the information score, the higher the correlation becomes between the BTS indexes and response accuracy. This result tells us that if the majority opinion can identify the correct answer the majority of the time, the information score actually hinders performance and thus should receive less weight in the formulation of BTS in order to improve its accuracy in identifying expert respondents.

8.4.3 *Does the BTS information or prediction score correlate with genuine knowledge?*

The results from our comparison of the correlation between the information and prediction scores and the BTS indexes of expertise revealed that while the prediction score correlated significantly with response accuracy, it did not outperform either index of expertise. In other words, the information and prediction scores alone are not better indicators of genuine knowledge than the BTS indexes of expertise.

8.5 *Future Directions*

The results from this study identified some of the alterations to the BTS equation that aids in the identification of correct answers and expert respondents. However, the current study only tested BTS’s ability to identify expert respondents in lower levels of learning (remembering, understanding and applying) or through the first phase of design (analysis). Therefore, the following chapters outline experiments developed to test BTS’s ability to identify expertise in the higher levels of learning (analysis, evaluation and creation) and later stages of design (idea generation and evaluation) and tested how alterations to the BTS algorithm can perhaps improve its performance.
Chapter 9

Experiment 5: BTS and Idea Generation

The previous experiment was designed to evaluate the Bayesian Truth Serum’s ability to predict genuine human factors knowledge as measured by response accuracy, test average and course grade. The results demonstrate that the BTS pooled index is a good indicator of genuine knowledge, but it does not outperform the index of conventional wisdom.

Recalling Benjamin Bloom’s Revised Taxonomy of Educational Objectives discussed in Chapter 2, we can conclude that the BTS assessment completed in Experiment 4 tested lower levels of the taxonomy, namely students’ ability to remember the human factors principles (remember), understand and apply the knowledge to rate the examples presented or their ability to successfully complete phase 1 of the design process (analysis). The aim of Experiments 5 and 6 is to test the higher levels of the taxonomy, namely the participant’s ability to evaluate design ideas and create new ideas, or their ability to successfully complete the later stages if the design process (idea generation and evaluation) for correlation with the BTS measures calculated in Experiment 4. In other words, how do the BTS indexes of expertise, which are intended to be used as a measure of knowledge, correlate with the highest levels of learning?
In the current study, we compare student’s ability to develop solutions to a human factors design problem and compare this ability to their BTS score calculated in Experiment 4. This study moves beyond the last experiment because we’re judging not only participants’ ability to identify high quality design examples, but their ability to produce them.

9.1 **Methodology**

This section describes the participants, the task, metrics used for comparison, the experimental design and the results.

9.1.1 **Participants**

Students were recruited for this study from the Human Factors Engineering Design Course at the University of Illinois. In all, 111 students with less than 1 year experience in Human Factors Design participated in this study. Of these 111 participants, 97 participants had also completed Experiment 4. Since we wanted to compare the BTS score from experiment 4 with the results from experiment 5, only these 97 students were used in the data analysis. Each student used the same 3 digit ID from experiment 4 to ensure anonymity and students participated on a voluntary basis.

The study involved two phases which took place over a two week time period with each phase taking approximately one hour to complete.

9.2 **Procedure Overview**

Experiment 5 is broken down into two phases: idea generation and idea critique. During the idea generation phase, participants were provided with a design problem and asked to develop
design solutions. During the idea critique phase, participants in the study rated the quality of the design ideas develop in phase 1 and provided a critique of the design idea. Participants did not receive a grade for this experiment but were asked to perform the task to the best of their ability. This section provides a detailed explanation of each of these phases.

**Phase 1: Idea Generation**

Upon arriving in the laboratory, participants read and signed an informed consent document. Once completed, the following instructions were provided:

“For today’s experiment we ask that you do the following:

**Generate (and sketch) as many** design ideas as possible for the design problem described below. The quality of the sketches bears no meaning, and as such, no time should be wasted on 'un-useful' sketching.

**Write comments** on each design such that an outsider could look at your idea and understand the concept being depicted. The quality of the comments is important as you will be receiving feedback on your idea. As such, an 'outsider' needs to be able to understand what your sketch is depicting without any additional information.

Finally, **number** each idea in the order you generate them.”

Once participants read the instructions, any questions were answered. Next, participants were given the design task and asked to develop solutions.

When developing the design task for this experiment, we wanted to select a problem that was both challenging to solve as well as familiar to the participants in this study so that our participants would be able to develop unique solutions in the hour timeframe. One problem on college campuses is the number of collisions and accidents that occur from pedestrians walking and talking on cell phones or listening to MP3 players (Nasar, Hecht, & Wener, 2008). For pedestrians, the use of these devices causes cognitive distractions, reduces situational awareness and increases unsafe behavior putting pedestrians that use these devices at a greater risk for
accidents. As most students own a mobile device and/or an MP3 player, they have the basic information necessary to understand the problem and develop solutions. In addition, there are many types of solutions to the problem, making it a great candidate for this study.

During the study, the following task description was read aloud to participants as they followed along on their own copy:

“There has been an increase in student accidents on campus in recent years from student’s texting and/or talking on cellphones or listening to music on an MP3 player while walking around campus. While using these devices, students become distracted, and can trip, fall or even run into something. In fact, in 2008 over 1,000 pedestrians visited emergency rooms due to accidents from using these devices while walking. There are reports of concussions, sprained ankles, broken appendages and even fatalities from these accidents. These numbers do not include the countless number of unreported incidents involving walking into something (i.e. a parked car) without an ER visit. This increase in accidents has been substantial on college campuses because of the number of students on campus and the increase of MP3 players and cellular usage (texting and talking), all of which are distracting.

As a Human Factors designer, you’ve been tasked with developing **AS MANY** solutions that would address all or some of the issues associated with this increased accident rate. These ideas can include either alterations to current technologies or the design of new technologies to help reduce these accidents.”

Once the problem description was read, any questions were answered. In addition, participants were reminded that the goal was to generate as many potential solutions to the problem as possible, focusing not on the quality of the drawings, but on the quality of the ideas. Participants were then given one hour to generate as many potential solutions to the problem as possible.

Once the hour had expired, participants were asked to select, by marking it with a sticker, what they felt was their ‘best idea’.

142
**Phase 2: Idea Critique**

Phase 2 occurred exactly one week following phase 1. The purpose of this phase was to provide a quality rating of the ideas developed in phase 1. Again, participants were not graded for this activity, but were asked to complete the task to the best of their ability. Participation in this study was a voluntary activity.

At the start of this study, participants read and signed an informed consent document. Once completed, the following instructions were provided:

“For this study, you will be asked to evaluate the ideas produced by other participants in IE 340 on a scale from 1 (poor human factors design) to 4 (excellent human factors design). In addition to rating the quality of the design in terms of HUMAN FACTORS PRINCIPLES, you will also be asked to provide a design critique. Here are some important things to remember when rating these design ideas

1. Consider the human factors principles discussed in lectures throughout the semester.
2. **DO NOT** evaluate the quality of the idea based on the designers sketching abilities, but instead your judgment should be based purely on the quality of the idea developed.
3. **DO NOT** judge the idea on its technical feasibility (i.e. can this idea actually be developed with existing technologies).
4. **DO** provide a critique based on the principles discussed in lecture.”

Participants were then handed a packet of 15 anonymized ideas that other participants had self-selected as their ‘best idea’ in phase 1. Each participant received a different set of 15 ideas in a randomized order, and participants did not rate their own idea. Participants were then directed to surveygizmo.com where an online survey was developed to help guide them through the critique process, see Figure 38.

Participants logged into the survey using their unique 3 digit ID provided in experiment 4. Next, participants were asked to provide the idea code present on each anonymized idea which was different than the participant id to insure anonymity. Participants were then instructed to provide a critique of the design idea based on the principles taught in class, i.e. identify both
strengths and weaknesses of the design, remove personal preferences and provide meaning-level feedback.

Figure 38: Survey used in Experiment 5, phase 2 to gather the participant's critiques and idea ratings.
Once the participant finished their idea critique they were instructed to rate the quality of the idea based on human factors principles on a scale from 1 (poor design idea) to 4 (excellent design idea). Finally, similar to in study 1, participants were asked not only to provide a personal answer when rating the design ideas, but also a prediction of the empirical frequencies that would endorse each response. In other words, they rated each idea on a scale from (1-4) and then predicted the proportion of other IE 340 students that would select each of the four responses. Participants were given 1 hour to complete as many of the 15 idea critiques as possible without compromising their critique quality.

9.2.1 *Experimental Design*

There were two main questions to be addressed in the second experiment:

(1) Can the BTS indexes of expertise better predict one’s ability to generate high quality design ideas over traditional measures such as course grade and test average?
(2) Do the pooled or individual BTS indexes of expertise correlate across experiments?

In order to answer question 1, an ‘idea score’ was calculated based on the peer ratings (1-poor design idea, 4-excellent design idea) for each idea:

\[
\text{idea score} = \frac{\sum \text{idea ratings}}{\text{total points possible}} = \frac{\sum \text{idea ratings}}{4 \times \text{number of idea ratings}}
\]

The idea score, in addition to the median response, was then tested for correlation with the BTS indexes of expertise as well as the participant’s course grade and average test grade using the two-tailed Pearson Correlation test in SPSS v. 18 with an \( \alpha \) of 0.05.

In order to answer question 2, we calculated the BTS expertise indexes based on the participants personal ratings of idea quality and predictions of the sample population that would endorse
each response captured in phase 2. Because a small sample population was used to rate each idea, between 7 and 10 participants rated each idea, steps 1-4 of the BTS algorithm were adjusted such that the participants own answer was excluded from the statistics that determined his or her score (Prelec, 2004). These BTS pooled and individual indexes were then tested for correlation with the indexes measured in Experiment 4.

9.3 Results

This section serves as a summary of the results from Experiment 5.

9.3.1 Do the BTS indexes of expertise correlate with idea quality

In Phase 1, each participant generated between 1 and 7 ideas (average 3.5). The best idea, self-selected by the participant, was then rated in phase 2 by 7 to 10 participants, average of 9.5, on a scale from 1 (poor human factors design) to 4 (excellent human factors design). Figure 39, Figure 40, and Figure 41 show the ideas that received the highest, intermediate and lowest idea scores, respectively. In order to determine if the BTS indexes of expertise could better predict the ability of respondents to create high quality design ideas over traditional measures like course grade and test average a Pearson two-tailed correlation test was performed see Table 8.
Figure 39: This idea received the highest idea score in Experiment 5.

- Transparent GPS sticky patch for any kind of music devices.
- GPS made by nanotechnology which minimizes the size and weight.
- By placing the patch on the back of the music devices, such as MP3, the GPS will turn on by using the electromagnetic wave from the MP3.
- It will alert the MP3 user by automatically pausing music, when the GPS senses that the user is close to any streets or intersections which the user has to cross while listening to music.
Figure 40: This idea received an intermediate idea score in Experiment 5.
Figure 41: This idea received the lowest idea score in Experiment 5.
The results showed that the pooled and individual index do not correlate significantly with the idea score or median idea rating. On the other hand, the student course grade did correlate significantly with the idea score ($r=0.19$, $p=0.05$) and the test average correlated significantly with the median idea rating ($r=0.22$, $p=0.03$).

<table>
<thead>
<tr>
<th></th>
<th>Median Idea Rating</th>
<th>Idea score</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Course Grade</strong></td>
<td><strong>Pearson Correlation</strong></td>
<td>.156</td>
</tr>
<tr>
<td></td>
<td><strong>Sig. (2-tailed)</strong></td>
<td>.127</td>
</tr>
<tr>
<td></td>
<td><strong>N</strong></td>
<td>97</td>
</tr>
<tr>
<td><strong>Test Average</strong></td>
<td><strong>Pearson Correlation</strong></td>
<td>.217</td>
</tr>
<tr>
<td></td>
<td><strong>Sig. (2-tailed)</strong></td>
<td>.032</td>
</tr>
<tr>
<td></td>
<td><strong>N</strong></td>
<td>97</td>
</tr>
<tr>
<td><strong>Individual Index</strong></td>
<td><strong>Pearson Correlation</strong></td>
<td>.063</td>
</tr>
<tr>
<td></td>
<td><strong>Sig. (2-tailed)</strong></td>
<td>.541</td>
</tr>
<tr>
<td></td>
<td><strong>N</strong></td>
<td>97</td>
</tr>
<tr>
<td><strong>Pooled Index</strong></td>
<td><strong>Pearson Correlation</strong></td>
<td>.045</td>
</tr>
<tr>
<td></td>
<td><strong>Sig. (2-tailed)</strong></td>
<td>.059</td>
</tr>
<tr>
<td></td>
<td><strong>N</strong></td>
<td>97</td>
</tr>
</tbody>
</table>

Table 8: Correlation test between idea score, traditional measures of design skill (course grade and test average) and the BTS score of each individual calculated in Experiment 4.

Although the correlation between the BTS indexes of expertise were not significant, we again hypothesized that we may be able improve the correlation between the idea score and the BTS indexes of expertise by performing the same weighting to step 2 of the BTS calculation that we performed in the previous chapter. In order to test this theory, we calculated the individual index and pooled index using 5 values of $d$ and tested these calculations for correlation with idea score, see Table 9. Unlike the results from Chapter 8, by changing the weight of the information score we do not see a significant change in the correlation between the BTS indexes of expertise and the idea score. What this means is the BTS indexes of expertise cannot be used to accurately predict one’s ideation abilities. However, the measures of course grade and test average can predict to some degree the student’s ideation abilities.
Table 9: The correlation between the BTS indexes of expertise and the idea score at varying levels of d, the weight of the information score in step 2 of the BTS calculation.

9.3.2 **Do BTS scores correlate across experiments**

In order to determine if the BTS indexes of expertise tabulated in Experiment 5, phase 2, mimicked those calculated in Experiment 4, a correlation test was performed, see Table 10. Our results show that the individual index correlates significantly across experiments (r=.224, p<0.05), but the pooled index does not.

Table 10: Correlation test between BTS expertise indexes calculated in Experiment 4 and Experiment 5, phase 2. The results show that the individual index correlates across experiments, but the pooled index does not.

9.4 **Summary and Discussion**

The goal of the current experiment was to determine BTS's ability to predict higher levels of Bloom's Taxonomy, the creation of ideas, and the reproducibility of our BTS results from experiment 4 when we have a smaller sample size and a different task. First, the discussion will focus on answering the experimental questions posed earlier in the chapter.
9.4.1 *Can the BTS indexes of expertise predict one's ability to generate high quality design ideas?*

In experiment 4 we found that the BTS indexes of expertise were able to identify expert respondents according to lower levels of Bloom's Taxonomy (remembering, understanding and applying). In experiment 5 we wanted to test if the same indexes were able to identify those individuals that had achieved the highest level of learning (creating).

The results from experiment 5 showed that neither BTS index correlated with the idea score or median idea rating. In fact, both indexes of expertise were far from being significantly correlated ($p>0.36$ for all correlations). Even when we weighted step 2 of the BTS score and removed all of the weight from the information score, the indexes did not show a significant relationship with the idea score. On the contrary, the traditional measures of course grade and test average did correlate significantly with the idea score and median idea rating.

This finding should not come as a surprise as the BTS score tabulated in experiment 4 was based on the lower levels of blooms taxonomy while we were trying to predict the highest level of the taxonomy. On the other hand, the course grade for IE 340, the class in which the assessment was performed, included not only homework, participation, and test grades, but also lab grades. These labs for this class included design activities, and thus the course grade included some measure of a student’s ideation abilities. Therefore, it is not surprising that the course grade would be a better indicator of this higher level of learning compared to the BTS indexes of expertise. What these results indicate is that the BTS indexes of expertise, calculated by rating design examples in terms of human factors principles, does not provide us with a better indication of design ideation abilities than the traditional measure of course grade.
9.4.2 *Do the pooled or individual BTS indexes of expertise correlate across experiments?*

In experiment 5 we calculated the BTS indexes of expertise for participants based on their ability to predict other IE 340 student’s ratings of idea quality. We hypothesized that those students that received high BTS scores in experiment 5 would also have received high BTS scores in experiment 4. Our results showed that the individual index correlated across experiments 4 and 5 while the pooled index did not. This could be attributed to the pooled index filtering out individual differences in prediction competence (i.e. long run calibrations of predictions). In other words, the pooled index is the average BTS score of the answer endorsed by the respondent averaged over all questions whereas the individual index is the average of the respondent averaged across all questions. Therefore, it makes sense that the individual index would correlate across experiments because it is a measure of the individual’s performance on each question and not an average of the group’s responses. What this means is the individual index may be a more rigorous metric compared to the individual index when predicting performance on similar assessments.

9.5 *Future Directions*

The results from this study indicated that the BTS indexes of expertise calculated in experiment 4 which were based on lower levels of learning are not good indicators of ideation abilities. However, the current study only tested the highest level of learning (creation) and thus there are two other higher levels of learning (analysis and evaluation) that may still be associated with the BTS indexes of expertise. Therefore, the following chapter will outline an experiment developed to test BTS’s ability to identify expertise in the remaining two levels of Bloom’s Taxonomy, or an individual’s ability to successfully complete the evaluation phase of design.
Chapter 10

Experiment 6: BTS and the Design Critique

Experiment 5 showed that the BTS indexes of expertise calculated in Experiment 4 failed to accurately predict ones' ability to create high quality design ideas. In other words, the indexes were not able to measure the highest level of Bloom’s taxonomy (application and synthesis of knowledge). Since the BTS calculations were based on lower levels of learning (remembering, understanding and applying), it is not surprising that these metrics did not correlate with the highest level of learning (creation). The purpose of the final experiment is to test the ability of the BTS indexes of expertise calculated in experiment 4’s ability to predict the next two highest levels of learning: analysis and evaluation.

In the current study, we compare the perceived utility of the student's design critiques from Experiment 5 with their BTS score from Experiment 4. This study moves beyond the first and second because we're judging the participants ability to properly evaluate the design idea and provide useful feedback in the form of a design critique.

10.1 Methodology

This section describes the participants, the task, metrics used for comparison, the experimental design and the results.
10.1.1 *Participants*

Students were recruited for this study from the Human Factors Design Course at the University of Illinois. In all, 111 students with less than 1 year experience in Human Factors Design participated in this study. Of these 111 participants, 94 participants had also completed Experiments 4 and 5. Since we wanted to compare the BTS score from experiment 4 with the results from experiment 6, only these 94 students were used in the data analysis. Each student used the same 3 digit ID from experiment 4 to ensure anonymity and students participated on a voluntary basis. The study took approximately one hour to complete.

10.2 *Procedure Overview*

Upon arriving in the laboratory, participants read and signed an informed consent document. Once completed, participants were provided with a packet that contained their original self-identified ‘best’ idea from experiment 5, phase 1, a copy of the 7-10 critiques of their ‘best idea’ (de-identified) from experiment 5, phase 2 and 5 blank sheets of paper, see Figure 42 for example critiques. The participants were then given the following instructions:

“For today’s experiment we ask that you do the following:

Read the critiques of your design idea in detail and think about how they can be used to improve your design.

Refine or redesign your idea based upon the feedback received from your peers. Come up with as many improvements as possible.

Write comments on each design such that an outsider could look at your idea and understand the concept being depicted.”

Once participants read the instructions, any questions were answered. Next, participants were given 30 minutes to refine their ideas based on the feedback they received from their peers.
Figure 42: Design idea develop by participant 097 in Experiment 5, phase 1 and critiques developed during Experiment 5, phase 2. These critiques were provided to participants in Experiment 6 to help them redevelop/refine their idea and rate the utility of the critiques for these purposes. The numbers on the right side represent how ‘useful’ participant 097 thought each critique was (4 being extremely useful).
Once the 30 minutes had expired, participants were directed to a survey at www.surveygizmo.com and were provided with the following instructions:

“In the following questions you will be asked to rate the critiques you received on the idea you developed. You will rate EACH critique on a 4 point scale from not at all useful (1) to very useful (4). This rating should be based on the value of the information inside the critique in terms of the feedback provided on Human Factors Principles and the value of the information for helping you refine your idea.”

The participant than inputted their 3 digit unique id from experiment 4 and rated each critique on its usefulness for helping the participant refine/ redevelop their idea, see Figure 43.

10.3 Experimental Design

The sixth experiment was developed to answer the following question: Can BTS better predict one’s ability to provide useful design critiques over traditional measures such as the participant’s course grade and test average?
In order to determine if the BTS indexes of expertise could better predict the ability of respondents to *analyze* and *evaluate* design ideas and provide useful design critiques, a correlation test was performed. The ratings of critique usefulness from experiment 5 were tabulated to calculate a ‘critique score’ for each critique from experiment 5 as follows:

$$\text{critique score} = \frac{\sum \text{critique ratings}}{\text{total points possible}} = \frac{\sum \text{critique ratings}}{4 \times \text{number of critiques}}$$

This critique score, in addition to the median response, was then tested for correlation with the BTS score from in Experiment 4 as well as the participant’s final course grade and average test grade using the Pearson two-tailed correlation test in SPSS v. 18 with an α of 0.05.

<table>
<thead>
<tr>
<th></th>
<th>Critique Score</th>
<th>Critique Median</th>
</tr>
</thead>
<tbody>
<tr>
<td>Course Grade</td>
<td>.031</td>
<td>.074</td>
</tr>
<tr>
<td>Sig. (2-tailed)</td>
<td>.769</td>
<td>.482</td>
</tr>
<tr>
<td>N</td>
<td>94</td>
<td>94</td>
</tr>
<tr>
<td>Exam Average</td>
<td>.061</td>
<td>.042</td>
</tr>
<tr>
<td>Sig. (2-tailed)</td>
<td>.562</td>
<td>.690</td>
</tr>
<tr>
<td>N</td>
<td>94</td>
<td>94</td>
</tr>
<tr>
<td>Individual Index</td>
<td>.229</td>
<td>.244</td>
</tr>
<tr>
<td>Pearson Correlation</td>
<td>.027</td>
<td>.018</td>
</tr>
<tr>
<td>Sig. (2-tailed)</td>
<td>.94</td>
<td>94</td>
</tr>
<tr>
<td>N</td>
<td>94</td>
<td>94</td>
</tr>
<tr>
<td>Pooled Index</td>
<td>.232</td>
<td>.224</td>
</tr>
<tr>
<td>Pearson Correlation</td>
<td>.025</td>
<td>.030</td>
</tr>
<tr>
<td>Sig. (2-tailed)</td>
<td>.94</td>
<td>94</td>
</tr>
<tr>
<td>N</td>
<td>94</td>
<td>94</td>
</tr>
</tbody>
</table>

Table 11: Correlation test between the Critique score from experiment 6, the BTS indexes of expertise calculated in Experiment 4 and the traditional measures of course grade and exam average.

### 10.4 Results: Does BTS correlate with critique quality

The results from experiment 6 show that neither the test average nor the course grade correlates significantly with the critique score or median critique rating, see Table 8. On the other hand, the BTS pooled and individual index calculated in Experiment 4 correlated significantly with both the critique score and critique median. What this means is the BTS score calculated in
Experiment 4 can predict one’s ability to generate useful design critiques while the traditional measures of course grade and exam average do not.

We hypothesized that we could improve the correlation between the critique score and the BTS indexes of expertise by performing the same weighting to step 2 of the BTS calculation that we performed in the previous chapters. In order to test this theory, we calculated the individual index and pooled index using 5 values of d (0, 0.25, 0.50, 0.75 and 1.0) and tested these calculations for correlation with response accuracy, see Table 12.

Similar to the results found in Chapter 8, we see an increase in the correlation between the pooled index and the critique score by reducing the weight of the information score in the BTS equation, see Table 12. In other words, as we remove weight from the information score, and thus add weight to the prediction score, we see a higher correlation between the pooled index and critique score. It is interesting to note that the correlation between the individual index and critique score is actually negatively affected by altering the value of d.

<table>
<thead>
<tr>
<th></th>
<th>d=0</th>
<th>d=0.25</th>
<th>d=0.50</th>
<th>d=0.75</th>
<th>d=1</th>
</tr>
</thead>
<tbody>
<tr>
<td>pooled index + critique score</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Pearson Correlation</td>
<td>.246</td>
<td>.245</td>
<td>.232</td>
<td>.190</td>
<td>.108</td>
</tr>
<tr>
<td>Sig. (2-tailed)</td>
<td>.017</td>
<td>.018</td>
<td>.025</td>
<td>.068</td>
<td>.305</td>
</tr>
<tr>
<td>N</td>
<td>94</td>
<td>94</td>
<td>94</td>
<td>94</td>
<td>94</td>
</tr>
<tr>
<td>individual index + critique score</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Pearson Correlation</td>
<td>.224</td>
<td>.226</td>
<td>.229</td>
<td>.224</td>
<td>.104</td>
</tr>
<tr>
<td>Sig. (2-tailed)</td>
<td>.031</td>
<td>.029</td>
<td>.027</td>
<td>.031</td>
<td>.320</td>
</tr>
<tr>
<td>N</td>
<td>94</td>
<td>94</td>
<td>94</td>
<td>94</td>
<td>94</td>
</tr>
</tbody>
</table>

Table 12: correlation between weighted BTS indexes of expertise and critique score
10.5 **Summary and Discussion**

The aim of Experiment 6 was to understand the utility of BTS for predicting individuals that can perform at two of the highest levels of learning (analyze and evaluate). To achieve this goal, we asked participants to rate the utility of the written design critiques received from other participants in Experiment 5. First, the discussion will focus on answering the experimental question posed earlier in the chapter.

10.5.1 *Can BTS better predict one’s ability to provide useful design critiques over traditional measures such as the participant’s course grade and test average?*

Experiment 5 results revealed that the BTS indexes of expertise were not good indicators of individuals that had achieved the highest level of learning (creating). It was hypothesized that the lack of relationship between these two components could be attributed to the types of skills being tested on the BTS test utilized in this dissertation; the test asked questions that required students to utilize lower levels of learning (remembering, understanding and applying) but we were trying to predict students that had achieved the highest level of learning (creating). Experiment 6 was developed to test if the BTS indexes developed in Experiment 4 could be used to predict individuals that had developed the next two highest tiers of learning (analyzing and evaluating).

The results from our study indicate that the BTS indexes of expertise were better indicators of critiquing abilities than the traditional measures of course grade and exam average. In fact, the traditional measures had no significant relationship with the critiquing abilities of the participants in our study. In other words, the traditional measures were not good indicators of students that had achieved higher levels of learning which is what most educators seek to test. We also found that by altering the weighting of the BTS algorithm we could improve the
correlation between the pooled index and the critique score. Similar to in Experiment 4, these findings suggest that the information score should receive less weight in tests of educational abilities because conventional wisdom is typically correct.

10.6 **Summary Part II**

From Experiments 4-6 we can conclude that the BTS algorithm is a better indicator of an individual’s ability to analyze and evaluate design ideas than the current measures of course grade and test average. Therefore, BTS could be used as a new assessment measure for these design skills. However, our results also indicate that BTS is not a good indicator of an individual’s ability to create high quality ideas. This could perhaps be linked to the types of questions utilized on the BTS test in Experiment 4 which required participants to analyze design examples as opposed to creating their own design ideas. Another possible explanation for BTS’s inability to predict individual’s ability to develop high quality ideas in our study is that students in this course may not have acquired this skill set as it is not a direct focus of the class. Thus, there may be little variation in the idea generations of the students in this course causing a lack of correlation with the BTS score from experiment 4. Future studies should be employed that test alterations to the BTS quiz with individuals that have been specifically trained in idea generation abilities to better validate, or invalidate, BTS as a measure of this skill level.
Example use is a common and important part of design practice and many researchers are beginning to explore how to support example-based ideation and assessment practices in design. Past research in this area has explored how examples are used during the idea generation process (Nickerson, 1999; Ward, 1994, 1995; Weisburg, 1995), how examples can inspire and constrain creative thinking (Jansson & Smith, 1991; Perttula & Sipila, 2007a; A. R. Purcell & Gero, 1996) and how one's ability to identify high quality examples relates to one's design expertise (N. Bonnardel, 2000; N. Bonnardel & Marmèche, 2004). Although prior research has identified examples as a cornerstone of the design process, there are still many questions on example-centric design methods that need to be addressed. For instance, how do designers collect examples during the design process, what types of examples do designers collect (constraining or inspiring) and how can example-based methods be extended for purposes outside of idea generation?

Therefore, the purpose this dissertation was two-fold. First, we sought to develop an in-depth understanding of example-centric ideation practices in order to develop implications for future design tools that better support these processes. Second, we sought to understand how examples
can be used as an assessment tool in engineering design. This chapter serves as a summary of these findings bringing to light the contributions of this dissertation, limitations of the research and future research directions.

11.1 Part I: Example-Based Ideation Practices

Three experiments were conducted in Part I of this dissertation to understand existing ideation and example usage practices. These studies focused on both engineering and creative design practices since examples are used in a variety of design disciplines and the generalizability of our results hinges on our understanding of example usage in the entire design community. These studies were based on interviewing, surveying and observing some of the best designers in the country. This section serves as a summary of these experiments.

11.1.1 Experiment Overview

The first experiment examined the idea generation techniques utilized in professional design practice and explored the utility of these methods throughout the design process. This was accomplished by conducting a semi-structured interview with eleven professional designers. The techniques and processes utilized were compared to those discussed in the literature base and the idea generation techniques were evaluated for their utility throughout the three phases of the design process.

In the second experiment we explored example use in a more detailed manner by observing and interviewing nine professional designers during a 90 minute session. In particular, we were interested in how and why examples were used during the design process. We were also
interested in how, and for what purposes, examples were used outside of the idea generation process.

Finally, in experiment 3 we explored how designers collect design examples on the Web and what difficulties they have during this process. This was accomplished by observing and interviewing designers during a 90 minute laboratory study. We were interested not only in the types of processes designers followed to collect these examples, but the types of examples they collect. These processes were compared across different design domains and different levels of expertise.

11.1.2 Review of major findings

In experiment 1, we developed a new design cycle and idea generation process model. The IR3 model of idea generation we developed emphasized the importance of non-distinct phases of idea generation (research, represent, refine, and inspiration) and the cycle within a cycle concept. This new model provides a deeper understanding of the idea generation process and also situates the idea generation techniques utilized during the design process in each of these phases. We also expanded the design literature base in this area by providing new insights on the idea generation techniques utilized in design practice by identifying 19 techniques. However, we not only asked designer to identify what techniques they used to generate ideas during the design process, but we also asked about what difficulties they had during idea generation. In our study, designers reported having difficulty finding experts to aid them in the ideation process, communicating with distributed groups, re-using previous design knowledge, and finding useful design examples to re-appropriate.
Experiment 2 built-off of experiment 1 by taking a more detailed look at the common but enigmatic practice of example use in design. By interviewing and observing some of the best designers in the country we were able to understand how designers search for, store and share design examples during the design process. We also built an understanding of why, and for what purposes, designers use examples throughout the different phases of design. From this study we were able to develop an in-depth understanding of the example-usage process and identify five implications for the development of future design support tools including altering Web-based search engines to reduce design fixation, improving the capture and visualization of search results, developing a better integration of physical and digital examples, helping designers recall why they selected and stored a particular example, and developing technologies that reduce the overhead of contributing personal examples to the design community.

In experiment 3 we wanted to understand what strategies designers use to retrieve examples from the Web and how (and when) they utilize the retrieved examples during the idea generation process. We also wanted to understand the types of examples collected during idea generation as previous research has shown that designers can become fixated on design examples that are familiar to the designer or similar to the design task. Both our qualitative and quantitative results showed differences in the utility of examples during the design process across design domains. For instance, industrial designers use examples frequently throughout the design process while Web designers only occasionally use examples during each of the phases of design. Graphic designers tend to collect examples early in the design process, but rarely utilize examples in the final phases of design. We also identified significant differences in the types and variety of examples collected during the idea generation process across both design domains and expertise. In particular, we found that Web designers collect more familiar examples than industrial designers putting them at a greater risk for becoming fixated during
the design process. We also found that expert designers collected more fixating examples than novice designers. From these results we were able to identify three implications for web-based design tools that aid in the serendipitous encounter of information, spawn creativity by automatic example re-composition. We also identified new design research directions in education and professional practice.

11.2 Summary Part II: Example-Based Assessment Practices

Part II of this dissertation was developed to test how examples could be used as an objective measure of student design skill in engineering design education. In order to accomplish this, three studies were conducted to understand how ratings of example quality relates to ones’ design knowledge, ability to generate high quality ideas and ability to provide high quality feedback in the form of a design critique. This was accomplished by using a novel application of the Bayesian Truth Serum (BTS), and was studied in a Human Factors Engineering Design course at the University of Illinois. This section serves as a summary of these experiments.

11.2.1 Experiment Overview

Experiment 4 was developed to test BTS’s ability to select the correct survey answer and predict a student’s ability to rate the quality of design examples based on human factors principles. A thirty question survey was developed to create a BTS score for each student. These thirty question comprised of pictorial examples in which students were asked to rate the examples quality on one of ten human factors design principles. For each of the thirty questions, the student was asked to provide a personal answer as well as a prediction of the empirical distribution of responses. With this information we calculated a BTS score for each answer on the survey and the BTS indexes of expertise for each participant. We also calculated the index of
conventional wisdom (% of times the participant answered with the majority) and the response accuracy of the participant (% of times they answered with the instructor). We then compared the answer selected by the BTS score for each question to the ‘correct’ answer to determine BTS’s accuracy. Next, we tested the correlation between response accuracy and the following measures: the BTS indexes of expertise, response accuracy, course grade and test average.

For experiment 5 we wanted to move beyond merely testing BTS’s ability to identify students that have mastered the skill of identifying high quality examples. Instead, we wanted to focus on how well the student’s BTS score (and traditional measures of course grade and test average) could predict the idea generation ability of the student. In order to accomplish this, we gave each student 1 hour to develop as many solutions as possible to a human factors engineering design problem. Their self-selected best idea was then rated based on the quality of the solution in terms of human factors design principles by other participants in the study. In addition to rating the idea, participants were also asked to write a critique of the design idea to help participants improve their original idea. These ratings of idea quality were then tested for correlation with the BTS scores calculated in Experiment 4 and the student’s course grade and test average.

Finally, in experiment 6 we wanted to see if the BTS indexes of expertise were able to accurately predict one’s ability to provide high quality feedback in the form of a design critique. In order to accomplish this, students were provided with the written critiques from experiment 5 and given 45 minutes to re-design their idea based on this feedback. At the conclusion of this time, students were asked to rate each design critique on its utility for redesigning their original idea. This rating was then tested for correlation with the BTS score calculated in Experiment 4, the student’s course grade and the student’s test average.
11.2.2 **Review of Major Findings**

In experiment 4 we found that the Bayesian Truth Serum was able to identify 23 of 30 correct answers on the survey, but it was not as accurate as conventional wisdom (majority response) which identified 28 correct answers. Because the majority opinion in our study selected most of the correct answers, we hypothesized that the ‘surprisingly common’ score tabulated in the BTS algorithm was actually harming BTS’s ability to identify the correct answer. To test this theory, we re-weighted step 2 of the BTS algorithm and found that by reducing the weight of the information score in BTS’s formulation (and thus increasing the weight of the prediction score), we were able to enhance its accuracy. In fact, when we removed all weight from the information score, BTS was able to identify 27 correct answers. In other words the ‘surprisingly common’ factor, or the information score, actually hindered BTS’s ability to correctly predict answers because the majority was correct most of the time.

The second goal of Experiment 4 was to determine if BTS could predict a participant’s genuine human factors design knowledge. In order to answer this question, we tested the BTS indexes of expertise for correlation with the measures of genuine human factors knowledge (response accuracy, course grade and test average) and found that both indexes correlated with response accuracy but they did not correlate with the student’s course grade or test average. In other words, those with higher BTS indexes of expertise were more accurate on the survey, but they did not necessarily have a higher course grade. This may mean that the BTS indexes of expertise are measuring different knowledge than that measured on tests in the course. In addition, although the correlation between the pooled index and response accuracy was noteworthy, the traditional measure of conventional wisdom was more strongly associated with response accuracy. This is due to the fact that the majority opinion was correct on 28 of 30 questions. By altering the weight of the surprisingly common score, however, we find we can increase the
correlation between the BTS indexes of expertise and response accuracy. This result tells us that if the consensus opinion can identify the correct answer the majority of the time, the information score may actually hinder performance and thus should receive less weight in the formulation of BTS in order to improve its accuracy in identifying expert respondents.

In experiment 5 we wanted to go beyond the research questions asked in experiment 4 and try and predict one’s ability to generate high quality design ideas with the BTS indexes of expertise which was based on ratings of design example quality. We performed correlation tests between the idea score generated in experiment 5 and the BTS indexes of expertise, course grade and test average. The results showed that neither BTS index correlated significantly with the idea score or median idea rating. Even when we weighted step 2 of the BTS score and removed all of the weight from the information score, the indexes did not show a significant relationship with the idea score. On the contrary, the traditional measures of course grade and test average did correlate significantly with the idea score and median idea rating. What these results indicate is that the BTS indexes of expertise, calculated by rating design examples in terms of human factors principles, does not provide us with a better indication of design ideation abilities than the traditional measure of course grade.

In the final experiment, experiment 6, we wanted to see if BTS could predict one’s ability to provide high quality feedback in the form of a design critique. We used the ratings of the design critiques developed in Experiment 5 and tested it for correlation with the BTS indexes of expertise, the participant’s course grade and the participants test average. The results from this study show that neither the test average nor the course grade correlates significantly with the critique score or median critique rating. On the other hand, the BTS pooled and individual index calculated in Experiment 4 did correlate significantly with both the critique score and critique median. What this means is the BTS score calculated in Experiment 4 can predict one’s ability to
generate useful design critiques while the traditional measures of course grade and exam average do not.

11.3 Implications

The implications for this dissertation centers around informing the two goals stated earlier in the chapter which include (1) developing an in-depth understanding of example-centric ideation practices in order to develop guidelines for computational tools that better support these methods and (2) understanding how examples can be used as an assessment tool in engineering design education. Implications for each of these two goals will be discussed individually.

11.3.1 Implications for Example-Based Ideation Methods

In the introduction we outlined how design has changed from a design from scratch to a design through synthesis environment where designers transform, combine or adapt elements of existing design examples to generate new ideas. Because design has become so example-centric, the creativity and originality of designers’ outputs is now largely dependent on their ability to select high quality and innovative examples for inspiration during the design process. However example-use had not been widely studied in the research community and thus, little was known about how designers collect examples during the design process and what types of examples designers use for inspiration. This was problematic both for the understanding of design activity as well as for the development of design tools that better supports these processes. This dissertation was developed to better understand these practices in order to develop guidelines for next generation design tools.
In particular, this dissertation outlined 19 techniques utilized by professional designers to generate ideas during the design process. It also highlighted the utility and frequency of these techniques throughout the design process. These findings add to the design literature base in terms of what types of techniques professional designers use, and what difficulties they have in these process. These findings also have implications for how to develop computational tools to support group collaboration, the elicitation of expertise, the re-use of design knowledge, and the support of actively and passively browsing for design examples. In addition, our results can help understand how to educate designers in idea generation by highlighting the types of techniques utilized in professional practice and the utility of these techniques for idea generation.

The results from this dissertation also revealed the importance of using the Web for example-finding activities. Our results show that the designers use the Web to find examples because the ‘randomness’ and resourcefulness of the Web allows them to make more connections and draw more inspiration than traditional sources like books and magazines. However, there were two main issues designers identified with using the Web for inspirational searching in our studies: it is either too random providing too many irrelevant sources or too specific not allowing for the discovery of inspirational content which can lead to design fixation. The results of this dissertation identified seven ways we could alter the Web to better support example finding including allowing for serendipitous encounters on the web, altering search results to prevent fixation and supporting actively and passively searching the Web for inspiration. This finding adds to the design and Web search literature base by identifying ways we can alter existing tools or develop new tools to better support example retrieval on the Web.

In addition to highlighting the importance of example use during the idea generation process, the results from this dissertation also brought light the importance of using design example throughout the design process as a communication, idea generation and evaluation tool. The
implication of this result is to inform the existing design literature that example usage needs to be supported throughout the design process. This result also identifies the need for more design research on example-centric methods outside of idea generation.

Finally, our results showed that the utility of examples varies greatly across design domains and certain design domains may be more likely to become fixated when using examples. This finding can be used to drive future investigations in this field and help inform the development of tools to support designers in the collection of less-fixating examples.

11.3.2 Implications for Example-Centric Design Tools

The introduction of this dissertation also identified the need for new objective assessments of student design knowledge in engineering design education. In particular, it highlighted the potential benefit of using ratings of example quality as a predictor of student competence in design because previous research as linked ones’ ability to identify high quality design examples to their design expertise. However, no study had researched example-ratings as a potential assessment tool. Thus, this dissertation sought to understand how examples could be used in design assessments through a novel application of the Bayesian Truth Serum.

Our results indicate that ratings of example quality can be used to predict ones design competence in terms of their general design knowledge and their ability to provide feedback in the form of a design critique. We also found by altering the calculation of the BTS’s ‘surprisingly common’ score we can increase the correlation between the BTS indexes of expertise and their design competence. This finding has several implications. First, this finding is important for design practice as it has identified a more accurate way of assessing student competence in design (as compared to the traditional measures of course grade and test average). Next, it
identifies how we can alter BTS for use in assessment practices; if the survey is developed such that conventional wisdom is able to select the appropriate survey answer the majority of the time, than less weight should be given to the surprisingly common score. Finally, this finding is useful not only for the way that we access student competence in design, but it could also be used in professional practice as a way to identify job candidates with more design knowledge.

Although our research showed that BTS and ratings of example qualities can accurately identify individuals who can identify high quality ideas and produce high quality feedback our results showed that BTS cannot accurately predict ones’ ability to generate high quality design ideas. The implication of this finding is that it identifies the need for future research to explore ways to adjust BTS to better identify this skill set, or identify new measures that more accurately assess a student’s ability to generate high quality ideas.

Finally, throughout Part II of this dissertation we utilized a novel application of the Bayesian Truth Serum. In Experiment 4, we tested how accurate the BTS score was in determining the correct survey answer. We found that although the survey performed better than chance, the original calculation of the Bayesian Truth Serum was not optimized for the current study because the majority of participants were able to identify the correct survey answer. In other words, the surprisingly common score calculated as part of the BTS hindered its ability to select the correct survey answer because majority opinion was correct on most survey questions. By re-weighing the surprisingly common answer we were able to greatly increase BTS’s ability to identify the correct answer. This finding has implications for how to alter the BTS for cases when majority opinion is often correct.
11.4 Limitations and Future Directions

There were several limitations to dissertation. First, in Part I we explored example use in a variety of domains in order to increase the generalizability of our results. However, experiment 4 revealed that there were significant differences in the utility of examples throughout the design process and the types of examples collected during the design process across design domains. Therefore, because of this difference, future work should explore each design domain separately in order to increase the utility of the findings for a particular design field.

In addition, the Bayesian Truth Serum quiz was developed to predict a student’s design knowledge in a human factors engineering course. The focus of this particular course is to produce graduates that can identify good and bad human factors designs (examples) and address issues. Therefore, the rating of example quality is a core part of this course. This means that the results of this study should generalize to other engineering courses that have similar goals, but future work should explore how this type of assessment generalizes to other engineering classes.

Finally, our results indicated that BTS is not a good indicator of an individual’s ability to create high quality ideas. This could perhaps again be linked to the types of questions utilized on the BTS test in Experiment 4 which required participants to analyze design examples as opposed to creating their own design ideas. Another possible explanation for BTS’s inability to predict individual’s ability to develop high quality ideas in our study is that students in this course may not have acquired this skill set as it is not a direct focus of the class. Thus, there may be little variation in the idea generations of the students in this course causing a lack of correlation with the BTS score from experiment 4. Future studies should be employed that test alterations to the BTS quiz with individuals that have been specifically trained in idea generation abilities to better validate, or invalidate, BTS as a measure of this skill level.
References


183
Appendix A: Bayesian Truth Serum Survey

The Bayesian Truth Serum survey developed to test participants in Experiment 4 of this dissertation can be found in the supplemental file named herring_appendix_a.pdf.
Appendix B: True/False Quiz Experiment 4

The True/False quiz developed to test participants knowledge of the human factors principles tested in Experiment 4 of this dissertation can be found in the supplemental file named herring_appendix_b.pdf.