EFFECTIVE INQUIRY for INNOVATIVE ENGINEERING DESIGN

by

Ozgur Eris

Preface by Larry Leifer

Springer Science+Business Media, LLC

EFFECTIVE INQUIRY FOR INNOVATIVE ENGINEERING DESIGN

EFFECTIVE INQUIRY FOR INNOVATIVE ENGINEERING DESIGN

by

Ozgur Eris, Ph.D.
Stanford University

Springer Science+Business Media, LLC

Library of Congress Cataloging-in-Publication Data

A C.I.P. Catalogue record for this book is available
from the Library of Congress.

Effective Inquiry for Innovative Engineering Design
by Ozgur Eris
Cover design by: Ergin Sargin
ISBN 978-1-4613-4729-3 ISBN 978-1-4419-8943-7 (eBook)
DOI 10.1007/978-1-4419-8943-7

Originally published by Kluwer Academic Publishers in 2004
Softcover reprint of the hardcover 1st edition 2004

Printed on acid-free paper.

Contents

Preface

Engineering Design is a question-driven process?

This is not a punctuation error. It is the essence of Eris's book. A declarative statement, a decision made, is actually a constellation of questions. Can there possibly be decisions without questions? Ozgur Eris has some striking answers. I see them as a breakthrough. You need to know about them.

Engineering Design is a question-driven process!

This insight was first inserted into my awareness by Professor John Arnold, founder of the Design Division of the Department of Mechanical Engineering at Stanford University in 1960. My subsequent experiences in Robotics, Mechatronics, Human-Machine Integration, Knowledge Management Systems, and New Media Design have individually and collectively confirmed the proposition. Looking beyond personal experience, it has become a mature "belief system" in Stanford's Design Education and Design Research community. While it is less well-appreciated elsewhere, it may be one of the distinguishing features of Stanford's unique role in the Silicon Valley and beyond.

Unfortunately, direct evidence of the "question-drive" has proven to be elusive. Eris's book, building on over 20 years of inquiry and a dozen PhD theses, finally brings together the evidence, a working taxonomic framework, and a well-reasoned argument for duality between questions and decisions. Together, they forge a new plateau in our understanding of the "effective inquiry" process in innovative engineering design. In operational form, we have a refreshingly new "Design Thinking" model that is empirically grounded, an advance in Design Research Methodology.

Absent evidence, an alternative view, one derived from the study of decision-making has taken hold and matured to become Design Decision Theory. In part, its utility rests on the fact that decisions are usually found in formal documents, and at least some related consequences can be traced in other document citations. The same cannot be said of questions, especially those posed during the informal, formative, pre-publication phase of design thinking that is rich in questioning behavior, but rarely recorded. Curiously, failure to record seems to extend to our memory of these events, hence contemporary digital recording technology played a key role in capturing and dissecting the phenomena.

If questioning is so important, why haven't you been reading more about it? If it is so prevalent, where and how does it express itself? Even if one suspects that it is important, how does one go about fostering one's own questioning performance, and that of others? Figure-1 suggests that you not imagine a straight and narrow "path ahead," but that you purposefully craft a divergent path that is more likely to corral the essence of the decision space and bring you to identify and decide upon, the best idea.

Figure 1. The optimal path ahead may not be straight.

A fine artist, Ergin Sargin, has captured the essence of our quest to understand the insight that engineering design is all about questioning. The decision lies at the center. We find the decision space and define the decision options by a spiraling path that is mapped by the questions we ask. There is little value, and high risk, in taking the straight and narrow path, so well represented by the decision maker's exclamation mark. No decision can be better than the options created through effective questioning. Eris's book brings you evidence to support this metaphor and guidelines for formulating good questions.

There are important practical consequences for, amongst others, engineering design, innovation management, discovery science, and meta-data design. Going beyond the big effects, there are also everyday implications for creative activity any time, any where with anyone.

Larry Leifer
Professor, Design Engineering
Stanford University

Chapter 1

INTRODUCTION

Designing is question intensive. Experienced designers treat inquiry as an influential cognitive mechanism in their thinking. However, our formal understanding of the specifics of that mechanism, and at a higher level, the role of question asking during designing, is limited. The research presented in this book explores the issue from both theoretical and empirical perspectives. The findings allow for the development of a question-centric design thinking model. The framework that forms the basis of the model characterizes the process of inquiry in design thinking at an operational level, relates that characterization to existing decision making theories by arguing for a duality between questions and decisions, and maps the proposed duality onto the broader context of the design process. The validity of the model is demonstrated empirically by the discovery of a correlation between the question asking processes of design teams and their performance.

This book not only articulates those insights for the reader who is curious to learn more about the role of question asking in design, but also demonstrates the uniqueness of *design thinking* by identifying a specific class of questions that are characteristic of design situations. My intention is for the reader to walk away with a heightened awareness of the power of questions, and to encourage him/her to apply the fundamental elements of the effective inquiry process outlined in the model in his/her own design practices.

In this introductory chapter, I will discuss my motivation for focusing on the subjects of inquiry and cognition within a design context, and outline the guiding research questions and the main constituents of the work.

1.1 Why Study Question Asking?

Prior to discussing my personal motivation for focusing on the process of inquiry in design, I would like to mention two external and broader factors that influenced my decision: the value system that is embedded in the research and teaching institution I have been a part of while formulating and conducting this research, Stanford University's Mechanical Engineering Design Division, and the information technology revolution that began in the early 1990s.

The pedagogical principals employed in design education at the Design Division are fundamentally based on the premise that design is a question-driven socio-technical activity. Graduate students in engineering design are repeatedly exposed to this premise through various methodologies while completing their coursework and prior to formulating their research. These methodologies communicate the significance of asking questions during semi-structured need finding, problem (re)definition and (re)framing, and conceptualization exercises. They are most effective when practiced in project based settings, and are rather intuitive and informal. Even though the informal nature of these methodologies makes it difficult to attribute them to specific individuals, I can easily reference the instruction I received from Leifer, Roth, Faste, and Adams as having influenced me to appreciate the value and relevance of question asking in design [Leifer 1994, Roth 1995, Faste 1995, Adams 1996] as well as having influenced related research that has been conducted within the community [Baya 1996, Mabogunje 1997].

The implications of the information technology boom of the 1990s for the field of design research have been significant in drawing attention to the topic of inquiry. The need for "knowledge systems" that would support practicing designers were recognized, and initial feasibility studies regarding their design and implementation were undertaken. These studies highlighted two problematic areas: identifying the relevant information to be captured and stored, and accessing and retrieving it. Inquiry was identified as one of the mechanisms through which these issues could be tackled. If such systems could mimic the information requests of actual designers—their information seeking questioning behavior—they would be more effective. Kuffner & Ullman's early work in this area, followed by Baya's, were influential [Kuffner 1990, Kuffner & Ullman 1991, Baya 1996]. More recently, Ullman summarized the "progress toward the development of the ideal mechanical engineering design support system" [Ullman 2002], and Marsh and Wallace identified question asking as a mechanism that facilitates information flow between expert and novice designers in industry [Marsh & Wallace 1995, Marsh 1997].

The subject of question asking behavior of design teams caught my attention as a potential research direction during a video interaction analysis session. Data for the analysis were collected during a two week design project carried out by graduate engineering design students whose goal was to design, prototype, and race a paper bicycle. During the analysis, I began to pay close attention to the questions raised in the interaction, and their effect on the design decisions that followed. Some questions seemed to have a strong effect on pivotal decisions, and others dissipated and had no discernable impact. In either case, questions and decisions struck me as being tightly coupled at a conceptual as well as a pragmatic level.

One way of exploring that connection was to identify all of the questions and decisions that occurred during the interaction, and construct a "question-decision map." The intent was to test if such a representation might be useful in confirming the existence of a connection, and discovering relationships between the nature and timing of the questions and the decisions they led to.

However, during my initial attempts to construct a map, I realized that our formal understanding of questions—as they occur in a design context—was not comprehensive and operational enough to allow me to study their relationship to other subjects such as decision making. It was necessary to know more about the nature of questions and to be able to formalize descriptors of their occurrence before they could be related to descriptors of other subjects. A review of the design research literature revealed insights that were limited to the application of information seeking questions in design knowledge systems (as discussed above), and in the architecture domain, among others paradigms, to a theoretical paradigm that frames designing as inquiry at an abstract level [Schon 1983, Gedenryd 1989].

Therefore, instead of focusing on question-decision maps, I decided to develop a comprehensive framework on the nature of questions occurring in design contexts, operationalize that framework, and attempt to validate it in a series of quasi-controlled laboratory experiments. It is important to note that differentiating between questions that are asked in design and non-design contexts has implications. I will list them here, and discuss them in depth in Chapter 2.

This research is based on two fundamental premises:

1. It is valid and useful to treat designing as a "way of thinking," and thus, as a specific type of cognition.
2. Question asking while designing is influential to the thinking of designers. It is related to the cognitive aspects of their problem solving, creativity, decision making, and learning processes, and consequently, to their overall performance.

1.2 Why Study Design Cognition?

For the most part, research in engineering is focused on understanding and predicting the behavior of innovative artificial (man-made) systems by way of studying the physical, chemical, and more recently, biological principles that govern them. In practice, the fundamental competency of engineers is seen to be their ability to understand, synthesize, and apply principles associated with the natural sciences in creating new technologies that ultimately result in new products.

There is no doubt that we, as engineers, benefit greatly from studying and applying such principles. However, as our knowledge of them has grown, it has become apparent that our personal involvement in the design process as human beings is also important, and that there is a need to understand the principles that govern *our* behavior as designers. While the scientific understanding new technologies are based on is constantly advancing, the discrepancy between our knowledge of those technologies, and knowledge of ourselves as designers, is growing. Bridging this gap by addressing the human dimension is now seen as an opportunity for increasing design performance in industry.

One of the most intriguing components of that human dimension is related to the thought processes we employ when we design; our thought processes—our cognition as designers—govern the behavior of the systems we design as much as the scientific principles we apply to create them. Therefore, it is relevant to be concerned with what design cognition is, and how it can be studied, taught, and improved.

It is not clear when the term "design cognition" was first used. In a keynote speech, Pahl presented a brief history of the collaboration between cognitive scientists and design engineers, and argued that the knowledge of technical systems was not sufficient in understanding the thought processes that led to the synthesis of designs, and that studying those thought processes was critical in improving the proposed design methodologies [Pahl 1997]. Recently, several Ph.D. dissertations have been published as explorations in design cognition [Dylla 1991, Fricke 1993, Dorst 1997, Mabogunje 1997, Gedenryd 1998, Brereton 1999], and different research groups have began to address the topic directly (Birkhofer, Gero, Lindeman, and Leifer to name a few). Also, there are at least two internationally recognized conference series that are centered on the topic: Design Thinking Research Symposium (DRTS), and the International Conference on Design Computing and Cognition (DCC). The growing interest suggests that design cognition is becoming a prevalent approach in design research, and supports the first premise outlined in the previous section.

1.3 Research Questions and Approach

The research presented in this book consists of theoretical and empirical dimensions. The two dimensions build on each other; the results of the exploration in one dimension feed into and influence the exploration in the other dimension. The research questions that guided me throughout those explorations are summarized in the following sections.

1.3.1 Theoretical Dimension: Characterization of Question Asking in Design

The theoretical dimension addresses the following research questions:

- How can the nature of questions that are posed by design teams be characterized and categorized at an operational level?
- Is there a relationship between question asking and decision making in design? If there is, is it possible and meaningful to develop a unified question-decision centric theory of design?
- Does the relationship between question asking and decision making—if it exists—influence design performance? What is a relevant framework for measuring design performance?

1.3.1.1 The Nature of Questions Asked while Designing

One way of studying the nature of questions that are asked while designing is to develop a comprehensive taxonomy of questions, and use it as a coding scheme to analyze the thinking of designers. When developing the taxonomy, various principles can be applied to differentiate between the types of questions. For the purposes of this research, I focused on two such differentiating principles that are related: conceptual meaning of questions, and a convergent-divergent thinking paradigm that is reflected in questions.

The first principle, the conceptual meaning of questions, has been articulated and used in the formulation of semantic question categories by Lehnert [Lehnert 1978]. Her approach will be discussed in detail in section 2.1. Prior to adopting her categories and/or constructing additional ones myself, I reviewed five other published taxonomies of questions. The second principle, a convergent-divergent thinking paradigm that is reflected in questions, is an outcome of my analysis of those taxonomies. It yields two meta-classes, which are made up of some of the question categories constructed through the application of the first principle.

The understanding embodied in these two principles resulted in the adoption of Lehnert's semantic categories, and in the formulation of divergent question categories. Together, the categories formed a comprehensive and operational taxonomy of questions that are asked while designing. The specifics of that framework will be discussed in Chapter 3.

1.3.1.2 Question-Decision Duality

As I mentioned in the beginning of this chapter, I perceived a strong conceptual link between questions and decisions while observing a series of design team meetings. Although I concluded that I needed to characterize questions asked by designers in a comprehensive fashion prior to attempting to formalize that link, I still perceived benefit in considering the issue on a philosophical level. The result was an analytical argument regarding the existence of a duality between questions and decisions.

The duality is based on the premise that it is imperative to ask questions in order to make decisions, and make decisions in order to ask questions. In section 2.2, this argument is presented in detail and illustrated with transcript segments from one of the design team meetings. Moreover, the findings of the empirical dimension allowed me to revisit and validate certain aspects of this relationship by allowing me to map it onto the design process. That mapping will be discussed in Chapter 8.

1.3.1.3 A Perspective on Design Performance

The recognition of design cognition as a topic in design research is advancing our understanding of design performance. Traditionally, when considering engineering design performance, researchers have been predominantly concerned with developing ways of evaluating the performance of the systems engineers design, and focused on the outcome of the design process, the product. The recent focus on the human dimension of designing, and on design cognition, has introduced another perspective for considering design performance, the *designer*.

These two viewpoints suggest the existence of two types of design phenomena that can be evaluated: what occurs during design activity, and what results from and persists after design activity. Naturally, the metrics for evaluating the performance associated with each phenomenon will differ. If one grounds himself/herself in design activity and takes it as the reference point, it is appropriate to treat activity-based metrics as being “internal,” and outcome-based metrics as being “external.”

As outlined in the second premise listed in the previous section, this research supposes the existence of a relationship between design cognition and performance. Since design cognition is a phenomenon internal to design activity, a framework for measuring internal design performance is required to study that relationship. When developing a framework in order to satisfy that requirement, I utilized the activity-outcome distinction in formulating a question-centric internal design performance metric. The specifics of that framework will be discussed in Chapter 4.

1.3.2 Empirical Dimension: Three Experiments

The empirical dimension of this research entails making a series of detailed observations in two distinct settings, and analyzing the data according to the frameworks developed in the theoretical dimension. The first setting was a real-life design project, and lent itself to ethnographic observation techniques. The second setting was a quasi-controlled laboratory experiment, and lent itself to video interaction analysis. The research conducted in these settings can be summarized in three progressive steps:

1. Detailed observation and analysis of a real-life design situation for hypothesis generation.
2. Design of a laboratory experiment to test the hypotheses.
3. Redesign of the pilot version of the experiment, and the execution of the final version.

The following are the guiding research questions associated with these steps:

- What hypotheses can be constructed regarding question asking in design?
- How can those hypotheses be tested? How should a design experiment be characterized in terms of its requirements? Is that characterization applicable to design experimentation in general?
- How should a design experiment be executed?

In taking each step, I was influenced by a design research methodology that has been used at the Stanford Center for Design Research for over 15 years. It advocates that the researcher should go beyond merely observing and describing design activity to constructing meaningful interventions to test the gained insights by iterating a cycle composed of three phases: observe, analyze, and intervene. The structure associated with each empirical step is outlined in the following sections.

1.3.2.1 Hypothesis Generation in the Field

The first research setting, a real-life design project, enabled me to freely observe a design situation where a team of graduate engineering design students designed, prototyped, and raced a paper bicycle. A colleague and I "shadowed" the design team, videotaping the nine design meetings the team held over a period of two weeks.

During those observations, I paid close attention to the questions raised in the interactions, considered potential relationships between question asking and decision making, and began to regard question asking while designing as a process. Most of the research questions outlined in the theoretical dimension of this work stem from those initial observations and conceptualizations. A detailed discussion of those insights, and their transformation into testable hypotheses is provided in Chapter 4.

1.3.2.2 Characterizing and Designing a "Design" Experiment

The second empirical step is the design of a laboratory experiment. I identified seven design requirements under three experimental design criteria that needed to be satisfied for the experiment to test the hypotheses. The framework for categorizing questions (as outlined in the synopsis of the theoretical dimension in section 1.3.1.1), the hypotheses, and experimental considerations specific to design research served as natural design criteria.

The nature of the requirements, and the specifications for meeting them, are discussed in detail in Chapter 5. The requirements under the first two criterion, question categorization and hypotheses testing, are specific to this research. However, I would like to stress that the third design criterion is relevant, and even necessary, for design research in general as it tackles the broader issue of what constitutes an "experiment" in a design context. The requirements for the third criterion address the need to simulate the inherent complexity of designing by:

1. Favoring quasi-control as opposed to full-control when inserting control elements into the design scenario used during the experiment.
2. Promoting designing as opposed to problem solving in the experiment.
3. If multiple hypotheses are to be tested, advocating that they be tested in a single experiment.

The specifications that satisfy the requirements under all three criteria are discussed in the latter sections of Chapter 5. And finally, a known design scenario—the bodiometer design exercise—that embodies the specifications was identified, described, and modified. In the exercise, designers are asked to design and prototype a measurement device, which can be moved along human body contours to measure their length.

1.3.2.3 Redesign of the Pilot Experiment: The Definition of a “Good” Question

The third empirical step aims to augment the hypotheses, and ensure that the design exercise did indeed satisfy the requirements.

I conducted two pilot sessions of the experiment with six graduate mechanical engineering design students. The pilot runs proved to be very effective in achieving both goals. They resulted in changes to the structure of the design exercise and the design performance framework. Although most of those changes were minor individual adjustments, their combined contribution to meeting the requirements was significant. For example, observing a need to increase the duration of the exercise by 30 minutes during the pilot runs provided the teams in the final runs enough time to complete the number of design iterations they needed, which meant that the exercise was more realistic.

The pilot runs also allowed me to reflect on the relevance and validity of my hypotheses, and to refine them as necessary. They prompted me to consider what a “good” question might be in a design context, and to incorporate its characterization into one of the existing hypotheses. I also perceived the need to construct a new hypothesis when I considered the consequences of a “good” question as opposed to its characterization. After revisiting my observations of the paper bicycle design team, I postulated that good questions are associated with, and followed by, conceptual leaps, or discoveries.

I then conducted the redesigned version of the experiment with 36 graduate mechanical engineering design students working in 12 teams, analyzed the data according to the two theoretical frameworks, and tested the validity of the hypotheses. A detailed discussion on the redesign of the experiment and the modification of the hypotheses is provided in Chapter 6. The analysis of the data collected during the redesigned experiment is presented in Chapter 7.

Finally, a question-centric design thinking model is synthesized from the theoretical and empirical findings and presented in Chapter 8.

Chapter 2

QUESTION ASKING: A FUNDAMENTAL DIMENSION IN DESIGN THINKING

As mentioned in the introduction, this work operates under two premises:

1. It is valid and useful to frame designing as a "way of thinking", and thus, as a specific type of cognition.
2. Question asking while designing is influential to the cognition of designers. It is related to the cognitive aspects of their problem solving, creativity, decision making, and learning processes, and, consequently, to their overall performance.

These premises have two major implications. The first implication is that studying design cognition is a distinct and relevant approach to design research. The second implication is that treating decision making as the fundamental cognitive mechanism driving design performance—a prominent position within the field—requires further consideration.

This chapter consists of three parts. The first two parts, sections 2.1 and 2.2, stem from my motivation to put those implications into perspective. Section 2.1 deals with the first implication, and entails reviewing the design research field by categorizing the current research areas into four topics, and positioning design cognition within them. Section 2.2 deals with the second implication, and entails focusing on design cognition by proposing and considering relationships between two fundamental cognitive mechanisms in designing, decision making and question asking.

The third part, section 2.3, is a review of published taxonomies of questions. It represents my initial exploration on the nature of questions, and constitutes the first step in developing a coding scheme that can be used to analyze the question asking behavior of designers.

2.1 Contemporary Topics in Design Research

In the next four sections, I put the first implication listed at the beginning of this chapter into perspective by discussing the contemporary topics in design research and positioning design cognition within them. I classify the topics into four categories: design processes, social theories, design information, and design cognition.

After an initial consideration, one might argue that the four categories I propose overlap to the degree that they lack meaning. The categories are indeed strongly related. Nevertheless, I see them as being defined by well-pronounced differentiations within the field, strongly reflected in the motivations and products of distinct groups of researchers. On the other hand, I believe that the strong relationships, and even overlaps, between the categories can and should act as a basis for informing researchers on missing knowledge within their domains. For example, most design information and knowledge systems lack functionality that can be alleviated by utilizing the findings from the other three domains—it is poor practice to develop a design knowledge system that does not address the underlying social, cognitive, and process related elements.

2.1.1 Design Processes

Researchers studying design processes have traditionally been concerned with categorizing the workflow of designing by decomposing it to interrelated tasks. The goal is to construct formal design processes, and to extract methods for design practice from them.

Numerous influential design process models have been developed [Asimov 1962, Hubka 1982, Pugh 1986, Pahl & Beitz 1988, Ullman 1992, Otto & Wood 2001]. Since processes are abstractions, the principles for abstraction can and often do differ between these approaches. However, the basic tasks that make up processes are similar. What differentiates them are the specifics of the relationships between the tasks and procedures they embody.

In a representative model of the design process, tasks and procedures are outlined in the form of a flow chart [Hubka 1982]. Arrows between design tasks signify conceptual, logistical, and temporal relationships. Arrows pointing back at previously executed tasks identify iteration procedures and address the recursive nature of designing. A similar design process model developed by Pahl and Beitz is especially significant [Pahl & Beitz 1988]. Since its introduction, it has been recognized as an official standard in German industry, and been widely applied in the design of new products.

The tasks that serve as the basic elements in these two models are indeed similar; both processes are composed of tasks related to the generation and characterization of design requirements, concepts, representations, and specifications. However, they propose somewhat different procedures for executing them.

Design process models can be applied and practiced in two domains: product development institutions, and individual or small groups of designers. For institutions, design processes constitute directly applicable methods that can be used to structure product development projects. They also constitute frameworks for organizing human and physical resources; a group of people and space are associated with each task, i.e. requirements engineers, release engineers, test engineers, concept development laboratories, testing facilities, manufacturing plants, etc. In other words, in institutional settings, design processes have direct social and physical manifestations.

For an individual or a small group of designers, design processes constitute methods that can be internalized and practiced while designing. It is reasonable to assume that they influence the way designers think (this relationship will be discussed in detail in section 4.4.4). In order to test this assumption, it is necessary to observe how designers communicate and act since it is difficult to directly observe how they think. In other words, design processes do not necessarily have physical manifestations in the practices of individual designers, but can be assumed to influence their thinking.

2.1.2 Social Theories of Design

Social theories of design are essentially constructivist approaches. Researchers who are interested in developing social theories aim to describe design activity by observing, analyzing, and reconstructing the interactions of the involved parties. They primarily focus on the social elements of designing (the effects of the social relationships between the participants of the design activity on the activity itself and its outcomes) rather than the social implications of designs (the effects of the outcomes of the activity on broader social contexts such as society).

Cuff's research has been influential as a pioneering exploration in this domain [Cuff 1982]. Her work focused on the negotiation that takes place between architects and clients in architectural design practice, and challenged the myth of the architect as the driving force. She argued that, in practice, influence is "diffused" across all participants, including clients, and that qualities such as ambiguity, unexpected outcomes, and open-endedness are inherent elements of designing. Cuff concluded by stating that the final design emerges out of the interaction of the participants.

Bucciarelli studied two engineering design projects in industry by using ethnographic methods [Bucciarelli 1988, 1994]. The main premise of his study is consistent with Cuff's conclusion: design is a social process. Bucciarelli acknowledged the pivotal role of social interaction in design, and went further by stating:

> "Different participants think about the work on design in quite different ways. They do not share fully congruent internal representations of the design."

He built on that observation to propose the existence of "object worlds," which are "worlds of technical specializations, with their own dialects, systems of symbols, metaphors and models, instruments and craft sensitivities." In essence, he argued that each participant possesses an engraved set of technical values and representations, which act as a filter during design team interactions. For example, a structural engineer will relate to a design project by focusing on the strength of *the design* whereas a manufacturing engineer will do so by focusing on its manufacturability. Although they are working on the same design, their mindsets govern their viewpoints, and their perceptions of the design differ. Based on that observation, Bucciarelli argued that the resulting design is not simply a summation, but rather, an intersection, of the products of those viewpoints.

Minneman studied an engineering design team engaged in a series of design exercises during a workshop [Minneman 1991]. He advocated the need to go beyond mere observation, to intervention, in order to test gained insights. He reemphasized Cuff's and Bucciarelli's views on the role of ambiguity and negotiation—that they are inherent to designing and constitute a condition and a mechanism for understanding and structuring design activity. In his own words, Minneman's findings have the following implications:

- "Those insights [on the role of ambiguity and negotiation] shift the focus of group design support onto communication systems."
- "Design education should be refocused on teaching designers to better function in group situations."
- "Design management must encourage designers to work together."

The synergistic contributions of these three studies encouraged further interdisciplinary approaches to design research by demonstrating value in the use of cross-disciplinary analysis frameworks and methods to understand engineering design practices.

2.1.3 Design Information

Researchers interested in understanding the generation, capture and sharing of design information are strongly influenced by the recent developments in information technology. Although the term "information" is not explicitly defined in most of the publications in this field [Eris 1999], there seems to be an informal understanding of what it represents. That understanding can be made explicit with the following statement: design information is the content of communication generated while designing which needs to be contextualized in order to gain meaning.

The researchers' treatment of information leads me to associate information with communication in this definition. There seems to be a similarity in the usage of the word information[1], suggesting that, in a design context, all information is created with the intent of communication—if not right away, sometime in the future. The usage also leads me to view information as lacking any specific meaning; the communication needs to be interpreted for it to be assigned meaning, in which case it might be more appropriate to call it knowledge.

The findings of design information related research can be implemented in software tools that support information communication, capture, and reuse. The requirements for such systems are commonly based on empirical findings on the information-handling behavior of designers.

Kuffner and Baya directly focused on the information-handling behavior of designers during conceptual design [Kuffner 1990, Baya 1996]. Kuffner's framework is based on the formulization of the information requests of designers. He paid special attention to "the design information required to answer questions about the design and to verify and refute conjectures about the design." He demonstrated that designers are interested in information other than that which is contained in traditional design documentation such as blueprints and specifications.

Baya used a similar approach, and in a preliminary study, explored the question asking behavior of designers in order to understand their information needs. He went one step further than Kuffner by incorporating his initial findings into the development of an information management tool, DEDAL. The deployment and assessment of DEDAL in design situations enabled him to obtain some key results regarding the information-handling

---

[1] For instance, the usage by McMahon and Wood [McMahon 1999, Wood 1999].

behavior of designers. He discovered that designers move between different types of information on an average of 13 seconds, and that they can simultaneously handle up to 40 concepts while they design.

In light of these findings, Yen argued that concept generation and development occur most frequently in informal media where capture tools are the weakest, and developed a software tool, RECALL, that captures tacit information generated in multimodal design activity [Yen 2000]. By deploying RECALL, he demonstrated that the capture and playback/analysis of tacit information during concept development reveal the rationale behind the decisions that were made.

Yang anticipated the growing role of electronic information in design activity, and aimed to enhance the collaboration among design teams by developing a software tool that improves the indexing and retrieval of design information [Yang 2000]. Similar to Yen, she perceived value in capturing and indexing design information while it is being generated. Making the analogy to a traditional engineering logbook, she qualified her tool as an "electronic notebook," and argued that it provides a "rich, unfiltered history of a design project."

Frankenberger took a different position; based on her observations of engineering design practice in industry, she argued that it is necessary to study the information-handling behavior of designers in the context of the design situations they are in [Frankenberger 1999]. She distinguished between routine work and critical situations, and reported that designers contact their colleagues for information in nearly 90% of the critical situations. This finding is strongly echoed in Marsh's research [Marsh & Wallace 1995, Marsh 1997]. Frankenberger argued that the information needs of designers can be adequately supported by software tools only during routine work, and that during critical situations, social interaction cannot and should not be substituted for.

2.1.4 Design Cognition

The topic the research presented in this book falls under, design cognition, involves the study of the thought processes designers experience while they design. It might be appropriate to refer to these thought processes as design thinking; since cognition can be defined as "the act of knowing"[2] it is plausible to treat design cognition as being synonymous with design thinking.

---

[2] As defined in the Longman Contemporary Dictionary of English.

Research in design cognition is primarily focused on the individual designer. This attribute differentiates design cognition from the other design research topics discussed in the previous sections as they entail studying phenomena that are external to the individual designer, i.e. design tasks and procedures, information flow, social interaction. That is not to say that, in design cognition research, the individual designer is treated as an isolated entity whose internal mechanisms have little connection with other designers or the environment. On the contrary, studying such connections constitute a promising methodology for discovering what is taking place "inside" the mind of the individual designer. Brereton's work is a good example of this approach, where she treated the interactions between designers and hardware as elements of "distributed cognition" [Brereton 1999], and used them to explore the cognitive development and learning processes of individual designers.

Research in design cognition often entails the application of theories and methodologies developed in cognitive science to explain and model design activity. Lehnert, an artificial intelligence researcher, wrote [Lehnert 1978]:

> "Among scientists interested in cognition, there is no general agreement on how it can be best studied. Cognitive science is therefore characterized as an interdisciplinary area, to which contributions may be made by either computer scientists or psychologists. This may seem surprising at first, since computer science and psychology are not commonly considered strongly related fields of interest. Once one understands exactly how a computer scientist and a psychologist go about studying cognitive phenomena, however, the connection is less mysterious."

She then outlined the research methodologies of psychologists and computer scientists, compared them, and concluded that their frameworks are analogous—apart from psychologists choosing to conduct experiments and computer scientists choosing to write programs. Her point is that both are useful paradigms for testing educated guesses. The two paradigms are complementary since some cognitive behavior can be studied more effectively with experiments, and others with computer programs.

Lehnert's view still holds true. The distinction she has made between the experimental and computational research methods for studying cognition is visible in current design research: some design researchers study design cognition by programming and learning from computational models of designer behavior [Gero 1985], and others study it by conducting experiments that involve designers and simulate realistic design situations [Cross, Christiaans, Dorst 1996].

Theoretical methods, where the researcher relies primarily on analytical tools and anecdotal evidence in order to understand the cognition of designers, constitute a third research method. In the absence of repeatable research procedures, theoretical methods yield findings that are more subjective when compared to the findings reached through the other two methods. A representative example is Schon's influential work, The Reflective Practitioner, where he proposed a framework that describes the "professional artistry" of the individual designer. This professional artistry consists of five elements: knowing in action, reflection in action, conversation with the situation, reflecting on the situation, and reflective conversation with the situation [Schon 1983].

2.2 The Question-Decision Duality

Within the design cognition domain, much has been published on the roles of learning, knowledge representation, problem solving, and decision making in designing. These subjects have also been studied in other fields. In many cases, the contribution of design researchers has been the application of those understandings to describing and modeling design activity. However, as mentioned in the previous chapter, a fundamental cognitive dimension, question asking, has received limited attention. This is possibly related to the absence of a process-oriented theory of question asking that can be operationalized.

Therefore, in this section, I set out to demonstrate the significance of question asking as a cognitive mechanism in designing. I intend to accomplish this by supporting the validity of the implication of the second premise of this research listed at the beginning of this chapter (treating decision making as the fundamental cognitive mechanism that drives design performance requires further consideration) by reviewing decision-centric views in design research, and arguing for an inherent duality between questions and decisions.

2.2.1 Decision-centric views of Design Thinking

Several decision-centric design thinking frameworks have been proposed [Dieter 1983, Radford & Gero 1985, Rowe 1987, Pugh 1990, 1996, Hazelrigg 1999, Otto & Wood 2001]. The common underlying concept in these frameworks is to consider, represent, and model design thinking as a decision making process, and at some level, associate the quality of design decisions with design performance. A common motivation is to address the need for a rational design concept selection methodology.

Hazelrigg wrote:

> "In order to ensure that engineering design is conducted as a rational process producing the best possible results given the context of the activity, a mathematics of design is needed. It is possible to develop such a mathematics based on the recognition that engineering design is a decision-intensive process and adapting theories from other fields such as economics and decision theory."

He built on that argument by utilizing decision theories in constructing a set of axioms for designing, and in deriving two theorems. He illustrated this approach by considering a scenario, in which several people are guessing the number of M&Ms in a jar, which is meant to represent a competitive situation where designers are required to make a design decision in the presence of uncertainty. He first tackled the scenario through what he called the "conventional engineering approach," which entails modeling the volumes of the jar and the individual M&Ms and relating them to each other. He then tackled it by applying his theorems in producing a statistical model, which accounts for uncertainty, risk, information, preferences, and external factors such as competition (elements of Game Theory). His model resulted in a number of decisions, only one of which he computed as being optimal.

He then compared the conventional engineering approach with his, and concluded that his axiomatic approach yielded a more accurate representation, and produced results with a higher probability of winning. In his closing words, he remarked that "all engineering design is a matter of decision making under uncertainty and risk."

Radford and Gero also articulated a decision-centric view [Radford & Gero, 1985]. Their goal was similar to Hazelrigg's as both were interested in constructing mathematical models of designing. However, the approaches differ when the nature of the models is considered; Radford and Gero explored a deterministic model and accounted for dealing with ambiguity through optimization, whereas Hazelrigg advocated a probabilistic model which has elements of ambiguity already built in.

Radford and Gero began by acknowledging that different paradigms—numerical and qualitative—exist for understanding design activity, and provided their rationale for focusing on design decisions:

> "As a starting point we shall take the premise that the essential feature of design is the existence of goals—however ill-defined those goals—which

> makes the process purposeful and necessitates decisions about the best way to achieve those goals."

They then considered the relationship between design decisions and the performance of the solutions they led to:

> "The exploration of the relationships between design decisions and solution performances is fundamental to design—a process of predicting the performance consequences of design decisions and postulating the decisions which will lead to desired performance resultants."

Within this framework, they treated optimization as a method for "introducing goal-seeking directly into the process."

Dieter's approach was more pragmatic; he was directly concerned with design practice. He demonstrated the relevance of the application of existing decision-centric views in evaluating and choosing between alternative design concepts [Dieter 1983]. After briefly discussing decision making under risk and uncertainty, he illustrated the construction of a decision matrix in order to determine the utility values—intrinsic worth of outcomes—associated with competing design concepts. His method is based on utility theory, which formalizes the development values in decision making, and is very similar to the widely used "Pugh selection chart" methodology [Pugh 1990].

Dieter then introduced probability theory, which assesses the states of knowledge, and combines them with elements from utility theory in demonstrating the application of decision trees to design concept selection.

The common premise of these frameworks is that designers are faced with critical decisions after generating concepts, which constitute different choices with different outcomes. Applying decision theory principles can improve their decision making processes by aiding them in choosing the most appropriate concept to satisfy a certain set of constraints, preferences, and goals. However, there are limitations to modeling designing *as* a decision making process as the design process is much broader in scope and there are other cognitive dimensions that drive design performance. Therefore, current decision-centric views would benefit from the consideration of potential relationships between decision making and other cognitive mechanisms used while designing. I will discuss this in detail in the next section.

2.2.2 Associating Question Asking and Decision Making: Two Interdependencies

Studying decision making as a rational process, and considering its role in designing is valuable. The value of studying decision making as a rational process does not need explicit qualification as it has been rigorously argued for in many different domains. As Howard remarks, decision analysis is related to "the systematic reasoning about human action," and it "stands on a foundation of hundreds years of philosophical and practical thought" [Howard 1988]. He states that the "resurgence of the field in modern times began with statistical decision theory and a new appreciation of the Bayesian viewpoint." He defines decision analysis as "a systematic procedure for transforming opaque decision problems into transparent decision problems by a sequence of transparent steps."

I outlined the role of decision making in designing in the previous section and argued for a need to consider the relationships between decision making and other cognitive mechanisms fundamental to design thinking. I believe the most effective way of addressing that need is to ground the motivation and context of decision-centric views of design in observations of design activity.

The approach mentioned in Chapter 1 is one way of achieving this grounding: identifying questions and decisions that occur in design team meetings, constructing "question-decision maps" based on that information, and analyzing the interplay between questions and decisions to understand how they influence each other. Although this work primarily focuses on question asking for the reasons outlined in Chapter 1, I perceive value in developing a conceptual understanding of the relationship between questions and decisions. Guided by my empirical findings on question asking, I reconsider and operationalize a part of that conceptualization in the broader context of the design process in Chapter 8.

When considering the utility of decision-centric approaches in design research and practice, especially decision trees that associate information and knowledge with a decision/design process, it is beneficial to expand the scope of the consideration from just the decision making tasks to the entire design cycle.

This can be accomplished by considering the following questions:

1. How did the decision-maker reach a position from which he/she could map his/her knowledge onto a decision tree?
2. How is reaching that position related to the decision making process, and more importantly, to the design process as a whole?

These questions do not receive sufficient consideration from design researchers who take decision-centric approaches. That can lead to treating the decision making process *as* the design process—an unsound analogy. On the other hand, decision theorists acknowledge these issues by recognizing that decision analysis can only be practiced after the position described in the first question is reached.

Howard asks, “Is decision analysis too narrow for the richness of the human decision?” He then argues that “framing” and “creating alternatives” should be addressed before decision analysis techniques are applied to ensure that “we are working on the right problem.” On framing, he states: “Framing is the most difficult part of the decision analysis process; it seems to require an understanding that is uniquely human. Framing poses the greatest challenge to the automation of decision analysis."

The tasks Howard identifies as being problematic, framing and creating alternatives, are inherent dimensions of designing. Design researchers have been attempting to formalize them for decades. Therefore, while design researchers have much to learn from decision theorists, decision theorists have much to learn from design researchers as well.

In light of this discussion, let us return to the first question that was posed, “How did the decision-maker reach a position from which he/she could map his/her knowledge onto a decision tree?” It can be answered by asking another question, and letting its answer point at a duality between questions and decisions: “How reversible is a decision making process?” In other words, “If one starts with a decision and works his/her way back through the cognitive events that led to that decision, what will he/she do when he/she reaches junctions in the decision tree that are associated with clusters of information and knowledge?”[3]

The answer I propose in this book is that one needs to consider the *questions* that made the acquisition or creation of those clusters of information and knowledge possible, and understand the question asking process of the decision-maker.

I will illustrate this view with a data segment from one of the experiments conducted in the empirical dimension of this research. In the experiments, teams of 3 graduate mechanical engineering students were asked to design and prototype a device that measures the length of body contours. In this specific excerpt, the team members are making a decision on how many gear reduction stages there should be between the sensor and the readout of the device in order to provide a meaningful measurement to the user (Transcript 2-1). In the far right column, the 14 questions and 1 decision that occur during the interaction are tagged sequentially.

---

[3] This specific formulation was introduced to me by Larry Leifer during a private discussion in 2000.

Transcript 2-1. Design team members A, B, and C are making a decision on the number of stages of gear reduction between the sensor and the readout so that their device provides a meaningful measurement to the user. In the far right column, the 14 questions and 1 decision that occur during the interaction are tagged sequentially.

Time	Sub	Utterance	Tag
04:13	A	So, what kind of gear reduction did we decided we needed?	**Q1**
04:18	C	So, 0.25 inches…	
04:22	B	the circumference is…	
04:25	C	7…4…5…	
04:27	B	Do we wanna know the circumference then?	**Q2**
04:32	C	Right, not the area.	
04:33	B	The circumference is 2 Pi R?	**Q3**
04:36	A	Yep. [team calculates circumference together]	
05:12	B	So we want something to only go around once?	**Q4**
05:17	C	Right, 50 revolutions.	
05:21	B	150?	**Q5**
05:24	C	Right. How many teeth are on these guys (gears)? This one has 5,6,7,8.	**Q6**
05:29	A	Or we can also do the belts. We can have rubber bands, yah.	
05:39	A	Can I borrow the ruler?	
05:42	B	It seems like there are…Oh, it says on them actually. 24.	
05:47	C	That's 3. 3 to 1.	
05:52	B	And we need 50 to 1?	**Q7**
05:54	C	Yep.	
06:03	A	This is about a quarter of an inch, three quarters of an inch. [measuring with ruler]	
06:08	C	So, we'd actually need 3 stages? Is that right?	**Q8**
06:16	B	3 times 3 to the 2 is 27…	
06:19	C	So that would still give us 2 revolutions.	
06:22	B	Yeah, we need at least 4 stages.	
06:30	C	That should be kind of hard to read, wouldn't it?	**Q9**
06:36	A	Well, maybe we can rotate around twice? I mean it's not hard to realize if it rotates around once, then we just need to aim for half of that. Do you know what I mean…maybe…	**Q10**
06:47	C	So, which one of you has the smaller hands?	**Q11**
06:49	A	I have the smaller, probably smaller. I have long fingers.	
06:54	B	What was, what were yours?	**Q12**
06:57	C	40 inches.	
06:58	B	40 inches…	
07:01	C	So, with the smaller hand if you go around, and if it's over 27 then it doesn't matter if it goes around more than once.	
07:09	A	I would say that after we could have it go…the indicator could rotate around twice and a little bit before it's hard to read. Do you know what I mean?	
07:21	C	Okay, 3 stages seems appropriate, right?	**Q13**
07:25	B	Yes.	**D1**
07:27	A	Is that assuming that we have a bunch of little gears though?	**Q14**
07:31	C	I'm kind of going under the assumption that we'll get about the same the gear ratio out of the rubber bands, too, since they're about the same size.	

The most striking observation is that all 14 questions are directly related to the decision the team is considering, and influence the three and a half minute process that leads the team toward a consensus by providing structure for the discussion and generating/uncovering the necessary information. (Several other questions, which lead to the concept of "gear reduction," precede this interaction and are not a part of the transcript segment.)

The decision process is initiated by A, who brings up the need to make a decision on the gear reduction mechanism in Q1. In Q4, B proposes to set the gear ratio so that a full rotation of the dial covers the whole measurement range. C performs the necessary calculations for that concept, and in Q8, asks others to consider the validity of his calculations, which leads B to think that they need 4 stages. In Q9, C considers the legibility of the dial, and asks others to interpret if the scale that would result from the gear ratio B is considering would be acceptable. A must have agreed with C's concern since she proposes a new dial concept—the dial rotating twice—in Q10. After the team considers that concept, C decides that 3 stages would be necessary if the dial rotates twice, and asks the others to assess her conclusion. B immediately agrees, and using 3 stages emerges as the decision. However, A is somewhat skeptical and challenges that decision in Q14 by questioning an assumption behind it. C addresses her concern, A does not object, so the consensus is reached and the decision is made. Q2, Q3, Q5, Q6, Q7, Q11, Q12, and Q13 influence the process by uncovering information and knowledge relevant to the formulation of Q4, Q8, Q9, Q10, and Q14 and D1.

This illustrates a strong relationship, a *duality*, between questions and decisions, which can be articulated with two axiomatic interdependencies:

1. Every question operates on decisions as premises since the questioner must make choices regarding the content, structure, timing, and communication of the question. Questions are formulated. From the questioner's perspective, there is no such thing as an unintentional question (even though questions might have unintentional and unanticipated consequences—that is irrelevant to the formulation of the question and the questioner's motivation). Therefore, the questioner is bound to make decisions when formulating questions.
2. Conversely, every decision operates on questions as premises since decision making entails dealing with choices—decisions are devoid of meaning if a there is a single choice. Thus, there must exist a minimum of two choices, which constitute options that need to be contemplated, defined, compared, and valued by the decision maker. Questioning is the enabling mechanism. Therefore, the decision maker is bound to question when making decisions.

From these interdependencies, it follows that the quality of the decisions a designer makes is coupled with the quality of the questions he/she asks, and that question asking and decision making should be given a similar degree of consideration as topics of study in design cognition. This understanding can form the basis of a new unified question-decision centric design theory, where decision making takes place *during* question asking, and vice versa.

2.3 Learning from Existing Taxonomies of Questions

In this section, I explore existing knowledge on the nature of questions. I intend to apply that knowledge in laying out the foundations of a theoretical framework that would serve as an analysis scheme for the empirical part of this research, which entails observing designers working in teams and analyzing their thinking. Taxonomies of questions are forms of knowledge regarding the nature of questions that are especially suitable for that role; categories of a taxonomy can constitute natural units of a coding scheme that can be used in observation and analysis.

Therefore, in the next four sections, I review six relevant frameworks from five different disciplines: philosophy [Aristotle], education [Dillon 1984], artificial intelligence [Lehnert 1978], cognitive psychology [Graesser 1994], and design research [Kuffner 1990, Baya 1992]. In the following sections, I will consider each framework independently. In Chapter 3, I will compare and augment them, and develop a coding scheme.

2.3.1 From Aristotle to the Modern Scientist: Review and Classification of Research Questions

Dillon, an education researcher, reviewed 12 schemes for categorizing research questions [Dillon 1984]. The schemes were published in the fields of education, philosophy, psychology, and history. His goal was to understand more about the "kinds of questions that may be posed for research." He stated that the utility of his approach can be viewed in three dimensions: understanding, practice, and pedagogics of inquiry.

He argued that the first dimension, understanding of inquiry, can take place at three different levels: the individual study, a corpus of studies, and the enterprise of research in a given field.

The second dimension, "practice of inquiry," entailed applying the understandings gained at the three levels of the first dimension to research practice; the design of the research study is the focus as opposed to the understanding of it. The hierarchical classification scheme can outline a procedure for the types of questions researchers want to and can ask.

The third dimension, pedagogics of inquiry, is the application of the understandings gained at the three levels of the first dimension in teaching. This can be effective in teaching students how to construct their own research questions.

Dillon's review of the 12 schemes yielded mixed results. He found that a significant portion of the taxonomies did not operate on specific and consistent differentiating principles. The principles used in forming the categories in most of the taxonomies were not made explicit by the authors, and examination of the taxonomies failed to reveal them. Therefore, Dillon argued that most of the published taxonomies have limited utility.

However, Dillon perceived significant value in Aristotle's approach. As Dillon pointed out, Aristotle opened Book II of Posterior Analytics by proposing, "The kinds of question we ask are as many as the kinds of things which we know," and proceeded to identify four kinds of questions:

"1) Whether the connexion of an attribute with a thing is a fact,
2) What is the reason of the connexion,
3) Whether a thing exists,
4) What is the nature of the thing."

As these four categories illustrate, Aristotle's fundamental premise was to assume that our knowledge resides in the questions we can ask and the answers we can provide. After introducing the categories, Aristotle suggested a relationship between them by claiming, "When we have ascertained the thing's existence, we inquire as to its nature. When we know the fact, we ask the reason." Dillon interpreted that relationship as a "sequence of inquiry," which is composed of the following progression: existence, essence, attribute, and cause.

Dillon then presented his own categorization scheme (Table 3-1, column 2), which he stated was based on "Aristotle's few, short, and encompassing propositions." His scheme distinguishes between kinds of questions according to the extent of knowledge about some phenomenon P entailed in the answer. It consists of three main orders that are representative of the sequence, or, rather, of the hierarchy, of questions proposed by Aristotle.

The first order categories describe the properties of a phenomenon. The second order categories describe comparative relationships between phenomena. The third order categories describe contingent relationships between phenomena.

In order to determine the comprehensiveness of his classification scheme, he first demonstrated that all of the categories contained in the other schemes correlate with the categories contained in his scheme, and then extracted 924 "research questions" found in a sample of nine education journals for coding.

He reported that his scheme accounted for 99% of the questions. He estimated the comprehensiveness of the other schemes by attributing the proportion of questions accounted by the corresponding categories of his own scheme[4]. Since none of the other schemes correlated with his scheme completely, that approach resulted in the comprehensiveness of the other schemes to be less than 99%. He reported Aristotle's scheme to be 89.1% comprehensive, and the other schemes to be 37%-83% comprehensive.

2.3.2 AI Scientist's Approach: A Taxonomy of Questions for the purpose of Computer Simulation of Question Answering

Lehnert's work was aimed at laying out the theoretical foundations of a computational model—an artificial intelligence—that can answer questions [Lehnert 1978]. The computational implementation of her model is called "QUALM." In her model, she treated answering of questions as a process that can be broken down into two parts: understanding the question, and finding an answer. The first part has to do with interpreting the question, the second with searching the memory of the artificial intelligence for the best answer. The first part of her approach required the development of a taxonomy of questions[5], and will be discussed here.

QUALM was based on Shank's theory of memory representation called "Conceptual Dependency" [Shank 1972]. In Lehnert's words:

> "Conceptual dependency is a representational system that encodes the meaning of sentences by decomposition into a small set of primitive actions. When sentences are identical in meaning, the Conceptual Dependency representations for those sentences are identical."

Conceptual dependency assumes that "cognitive memory processes operate on the meaning of sentences, and not on the lexical meaning of those sentences." In other words, the fundamental operational mechanisms of memory are thought to be solely dependent on the conceptual meaning of what is being memorized, and to be independent of their lexical expression. For instance, the questions "Did Mary sell John a book?" and "Did John buy a book from Mary?" have similar conceptual representations.

As Lehnert stated, a fundamental element of conceptual representations are "primitive actions." Conceptual dependency does not specify a finite set

---

[4] Dillon argued that an indirect approach for determining the comprehensiveness of the other schemes is valid since he has proved his scheme to be encompassing of the other schemes as well as nearly all of the research questions in the data set, and that a scheme by scheme test was not necessary.

[5] Lehnert's taxonomy was not reviewed by Dillon as Dillon's focus was on research questions.

of primitives. However, the primitives it specifies are meant to constitute a small set so that its strength as a representation system is preserved.

Another fundamental element of conceptual representations is "causal chains." They are used to establish causal relationships between the events described by primitive actions. For instance, when Mary falls and breaks her arm, gravity propelling Mary to the ground and Mary getting hurt constitute causally linked events, and the causal link is defined as "RESULT." The following are the six basic causal links in Conceptual Dependency:

RESULT: An event results in a state.
REASON: Links mental events to non-mental actions.
INITIATE: A state or event initiates a thought process.
ENABLE: A state enables an event.
LEADTO: Links two events such that the causal chain is not explicit.
CANCAUSE: Modified LEADTO link where unspecified causal chain expansion is left out of the causal chain.

Lehnert argued that the most important dimension of a question that needs to be interpreted for it to be understood and answered appropriately is its conceptual meaning. She also stressed that lexical categorizations differentiating between the so-called what, how, and why questions "do not constitute a comprehensive system and are not motivated by anything greater than a desire to have a few general descriptive devices." (The empirical part of this research independently arrives at evidence supporting her claim.)

Lehnert then proposed her conceptual question categories, which are based on semantic differences. She thought of the categories as "processing categories that are predicted by features of conceptual representation." The following are her 13 categories (the descriptions and examples are summarized from Lehnert's detailed discussion):

1. Causal Antecedent: The questioner wants to know the states or events that have in some way caused the concept in question. The causal link is LEADTO.
   Example: Why did the glass break?
2. Goal Orientation: The questioner wants to know the motives or goals behind an action (commonly referred to as the why-question). Goal orientation questions are a specific case of the causal antecedent questions in the sense that the reason behind the concept is mental. The causal link is REASON.
   Example: Why did John take the book?
3. Enablement: The questioner wants to know the act or the state that enabled the question concept. The causal link is ENABLE.

Example: What did John need in order to leave?

4. Causal Consequent: The questioner wants to know the concept or causal chain the question concept caused. The causal link is LEADTO. Example: What happened after John left?
5. Verification: The questioner wants to know the truth of an event.
   Example: Did John leave?
6. Disjunctive: Verification question with multiple concepts.
   Example: Was John or Mary here?
7. Instrumental/Procedural: The questioner wants to know the partially or totally missing instrument in the question concept.
   Example: How did John go to New York?
8. Concept Completion: The questioner wants to know the missing component in a specified event (commonly referred to as the fill-in-the-blank question).
   Example: What did Mary eat?
9. Expectational: The questioner wants to know the causal antecedent of an act that presumably did not occur (commonly referred to as the why-not question). The causal link is LEADTO.
   Example: Why didn't John go to New York?
10. Judgmental: The questioner wants to solicit a judgement from the answerer by requiring a projection of events rather than a strict recall of events.
    Example: What should John do to keep Mary from leaving?
11. Quantification: The questioner wants to know an amount.
    Example: How many people are here?
12. Feature Specification: The questioner wants to know some property of a given person or thing.
    Example: What breed of dog is Pluto?
13. Request: The questioner does not want to know anything, but wants a specific act to be performed.
    Example: Can you pass the salt?

2.3.3 Cognitive Psychologist's Approach: Considering the AI Taxonomy in the Context of Educational Goals

Graesser was interested in understanding the role of question asking in learning, and identifying mechanisms that generate questions [Graesser 1988, 1992, 1993, 1994].

He stated that even though education researchers and teachers seem to agree on the "virtues of being an inquisitive learner who actively exerts control over the materiel to be learned by asking questions," most students are not active but passive learners, "who do not impose themselves on anyone with a question." He pointed out that studies have shown that the questions students ask are "infrequent and unsophisticated," and "constitute approximately 1% of the questions in a classroom, at an average of one question per hour" [Dillon 1987, 1988; Flammer 1981; Kerry 1987]. The questions students ask tend to involve "the recall and interpretation of explicit material rather than questions that involve inferences, application, synthesis, and evaluation." Also, attempts in facilitating the asking of more questions by students have resulted in an increase in the number of unsophisticated questions. And finally, teachers do not fare much better in asking sophisticated questions as "less than 4% of the instructor generated questions are higher-level."

The taxonomy of questions Graesser presented was based on Lehnert's framework (see section 2.3.2). Graesser adopted Lehnert's 13 semantic categories, and added five new ones. The categories he introduced are: "Comparison" (which he states was investigated by Laurer & Peacock, 1990), "Definition," "Example," "Interpretation," and "Assertion." Graesser did not provide a discussion on how the additional categories relate to the principles of Lehnert's taxonomy.

Graesser used the modified framework to analyze the frequency and the type of the questions asked by students during a series of tutoring sessions for an undergraduate class on research methods [Graesser 1994]. He focused primarily on student questions as they "reflect active learning," and not on tutor questions.

He concluded that the frequency of the occurrence of a certain class of questions correlate positively with student learning ($R = 0.46$, $p < 0.05$ as measured by an examination score), and termed them "Deep Reasoning Questions," or "DRQs." DRQs consist of the following question categories: Instrumental/Procedural, Causal Antecedent, Causal Consequence, Goal Orientation, Enablement, and Expectational. He argued that DRQs "tap the steps and rationale in logical reasoning, problem solving procedures, plans, and causal sequences."

In order to generate a stronger argument for the correlation between DRQs and learning, Graesser considered DRQs in the context of Bloom's taxonomy of educational objectives in the cognitive domain [Bloom 1956]. In Bloom's taxonomy, educational goals are organized into six hierarchical categories. Accomplishing the higher level objectives requires the mastery of the lower ones. This principle is similar to Dillon's principle regarding progression in inquiry (see section 2.3.1). Graesser argued that DRQs map onto the higher level educational objectives, and therefore, are indicative of student learning.

He coded the student questions that were asked in the tutoring sessions according to Bloom's taxonomy, and tested for correlation between DRQs and the proportion of questions that are regarded as comparatively deep in Bloom's taxonomy (levels 2, 3, 4, 5, and 6). His analysis yielded strong correlation ($R = 0.64$, $p < 0.05$). He also reported correlation between the questions that are regarded as deep in Bloom's taxonomy and examination scores ($R = 0.35$, $p < 0.05$).

Graesser outlined other descriptive data that are relevant to the empirical dimension of this research. He reported that the students in the tutoring sessions generated 21.1 questions per hour, and the tutors generated 95.2 questions per hour (yielding a combined rate of 116.3 questions per hour for the student-tutor couple). This is very high compared to the 0.11-0.17 questions generated per hour in the classroom by each individual student (as reported by Dillon, Flammer, and Kerry). If only the DRQs are accounted for, the rates drop down to 4.6 questions per hour for students, and 15.2 for tutors (yielding a combined rate of 19.8 questions per hour). There is no data on the DRQ asking rates of students in classrooms.

2.3.4 Design Researcher's Approach: Two Taxonomies on the Information Needs and Handling of Designers

Kuffner and Baya developed question-based research frameworks that can be operationalized. Kuffner was interested in characterizing the information designers require to answer questions and verify or refute conjectures about the design [Kuffner 1990, 1991]. Baya was interested in the nature of design information reuse and the role of questions in the information handling of designers [Baya 1992, 1996].

Kuffner's framework illuminated the relationship between questions and conjectures. The main principle he used to differentiate[6] between the types of questions and conjectures is their verification attribute. If a conjecture is not followed with an immediate attempt at verification, it is called a "simple conjecture." If it is followed with an immediate attempt at verification, it is called a "conjecture with verification." Somewhat similarly, questions requiring only simple answers are called "verification questions," and questions requiring detailed answers are called "open questions." Each

---

[6] In this book, the definition of a differentiating principle is taken to be an explicit rule, or a system of rules, that are used as the basis for expanding a phenomenon and constructing categories under it. For instance, if physical appearance is used as a differentiating principle for categorizing people, eye color, height, and weight would constitute valid categories, whereas name would not since it cannot be constructed through the application of the differentiating principle.

question and conjecture is also categorized according to its "Topic," "Age of its topic," "Nature," "Confirmation" and "Validity." Topic is the "design object the questioner focuses on". Nature is dependent on the "type of information that the subject either seeks or presumes." Confirmation indicates if the question or conjecture is confirmed, and if so by whom or what. Validity "measures the accuracy of a conjecture."

Baya observed that "it is very natural for us to express our information needs in the form of questions," and treated questions as identifiers of the content and the importance of the information designers seek. His question-centric framework reflects this thinking; the design information categories are identical with the question categories.

Baya categorized a question according to its "Descriptor," "Subject class," "Criticality," and "Level of detail." Descriptor refers to "the character or nature of the information being sought." It is almost identical to the "nature" class in Kuffner's scheme. Subject is "the subject of the sentence or the clause representing the questions." It is similar to the "topic" class in Kuffner's scheme. Criticality reflects the "measure of the impact asking of the question had on the overall goal of accomplishing a design." Level of detail is the level of detail of the information in the answer to the question.

Baya used the taxonomy to analyze two design sessions where individual designers were asked to redesign a shock absorber. His findings served as a set of requirements for the development of DEDAL, a design information utility. While commenting on the differences between his and Kuffner's frameworks, Baya made a key observation by stating that the questioning behavior of designers is not random, and that they ask new questions after reflecting on information received in answer to other questions.

Even though this observation is rather information-centric—not all questions are asked to seek information—it is significant in the sense that it touches upon the notion of treating question asking as a process.

Chapter 3

DEVELOPMENT OF A TAXONOMY THAT IS COMPREHENSIVE OF THE QUESTIONS ASKED WHILE DESIGNING

I took the initial step in the development of a coding scheme that can be used to analyze the types of questions asked by design teams by reviewing six taxonomies of questions in section 2.3. In this chapter, I first consider the comprehensiveness of those taxonomies, and then augment them. More specifically, my goals are to:

1. Discuss the appropriateness of treating the principles and question categories associated with the published taxonomies as analysis dimensions and units for studying the question asking behavior of designers.
2. Identify, if they exist, dimensions of the question asking behavior of designers that are not addressed by those principles.
3. Propose new principles and categories that will address any missing dimensions.

Fulfilling these goals would constitute the second step in the development of a coding scheme, and result in a theoretical framework.

In section 3.1, I provide the context for the observations that were instrumental in realizing these goals. In section 3.2, I discuss what constitutes a question in a design context, and arrive at a working definition. In section 3.3, I consider the comprehensiveness of the taxonomies reviewed in section 2.3, and identify a characteristic dimension of the question asking behavior of design teams that the taxonomies do not address. I then adopt one of the published taxonomies and extend it with the addition of five new question categories in order to make it more comprehensive. In section 3.4, I compare and contrast four of the reviewed taxonomies and my extensions.

3.1 Context for the Observations on the Nature of Questions Asked While Designing

Prior to discussing the comprehensiveness of the reviewed taxonomies, it is necessary to provide context for my reflection and evaluation. As mentioned in Chapter 1, the research presented in this book has empirical and theoretical dimensions. A critical component of the theoretical dimension is the development of a taxonomy of questions representative of the types of questions asked in design situations. The empirical dimension entails formulating hypotheses from field observations regarding the question asking behavior of designers, and testing those hypotheses by conducting laboratory experiments. The connection between the theoretical and empirical dimensions is the use of the taxonomy of questions to analyze the data collected during the experiments.

At first glance, these two dimensions might seem to be independent undertakings; however, they are corresponding endeavors. Although the start and end point of this research is a theoretical framework, my approach relies on establishing a dynamic dialogue between theory and empirical findings. The construction of a comprehensive and meaningful taxonomy is gradual and requires continuous reflection.

My process for maintaining that dynamic dialog is as follows: I begin with an existing taxonomy of questions synthesized from the contributions of researchers operating in different domains. I apply the taxonomy to the analysis of a design situation, and reflect on its appropriateness and utility in light of empirical data. This reflection allows me to make conceptual leaps in my understanding of questions. Each time I make a conceptual leap, I modify the taxonomy by refining existing categories and/or constructing new categories in order to incorporate the enhanced understanding. I then apply the augmented taxonomy to another design situation to generate more empirical data, and repeat the cycle. At the beginning of Chapter 4, I identify the three major steps that make up the empirical dimension of this research. Each step can be seen as one such cycle.

This cyclic approach produces a dilemma when it comes to presenting the findings that are embodied in the structure of the taxonomy. The gradual development of the understanding reflected in the taxonomy can be presented chronologically, or the final state of the taxonomy reflecting the most advanced understanding can be presented by itself. The chronological treatment is likely to be problematic and may confuse the reader by forcing the premature presentation of methodological discussions that are not directly related to the conceptual development of the question taxonomy. Those insights are best communicated separately. Therefore, I choose the second option, and present the most advanced understanding on the question categories in section 3.3.

The disadvantage of this approach is the absence of context for the discussion that I will present in this chapter. Naturally, the discussion will be much easier to interpret once the reader proceeds to read Chapters 4, 5, and 6. At this point, providing some background for the design situations I collected empirical data from might alleviate that limitation.

I observed two types of design situations. The first situation was a two week long real-life design project where a team of 4 graduate mechanical engineering students designed, prototyped, and raced a paper bicycle. The second situation was a set of 90 minute laboratory experiments where 14 teams of 3 graduate mechanical engineering students designed and prototyped a device that measures the length of body contours. The transcripts that I use to illustrate my arguments were extracted from the discourse of the teams who participated in the laboratory experiment.

3.2 Definition of a Question

Defining a question in a design context is challenging. Designers use a variety of communication mediums when engaged in design activity, and there are unique question posing opportunities associated with each medium. Gesturing [Tang 1991], interaction with hardware [Brereton 1999], sketching, speech, and written documentation are potential communication mediums. Apart from such mediums, which require the active participation of an actor in the formulation of a question, elements of the design environment can constitute embedded question asking mechanisms. For instance, the mere presence of a person or an object in the environment could constitute a question.

An explicit definition of a question in a design context was not provided within the question-centric design research frameworks reviewed in section 2.3.4. Also, the nature of questions was not considered on a comprehensive level. Instead, pragmatic aspects of question asking were addressed as the primary interest was in understanding information flow and processing during design activity.

However, as the frameworks reviewed in sections 2.3.1, 2.3.2, and 2.3.3 indicate, the topic of inquiry has received a much broader and comprehensive consideration in other disciplines. In the discussions of the reviewed frameworks, the authors often referenced questions as expressions in written or verbal language although their considerations were conceptual and independent of the medium questions were posed through. That tendency might stem from the fact that it is much more difficult to define and characterize questions communicated through mediums such as gesturing and sketching. In spoken and written language, there are many explicit signals that are built in such as grammar and punctuation.

This observation leads me to focus on the verbal exchanges between designers. I omit the written exchanges since, in this study, I focus on observing and analyzing design activity at the co-located team level, where written exchanges between designers are limited—if not nonexistent. Therefore, for the purposes of this research, I construct and utilize the following definition for a question:

> *In a design context, a question is a verbal utterance related to the design tasks at hand that demands an explicit verbal and/or nonverbal response.*

Even though this definition clearly limits the scope of my observations and their implications for reasons I mentioned earlier, I believe that it addresses one of the most common and influential modes of communication in group design activity, and, therefore, is a good starting point.

3.3 An Argument for the Search for the "Possible" and Its Characterization as Question Categories

When considering the comprehensiveness of the reviewed taxonomies I tested the appropriateness of treating their categories as analysis units for coding the questions that were asked in the two design settings. During the course of my analysis, I extracted over 2000 questions from the data collected during design meetings. When I used the reviewed taxonomies to categorize the questions, I could not categorize over 15% of them. Considering the nature of these questions and reflecting on why they were not represented in the reviewed taxonomies resulted in the identification of an overlooked principle.

The common premise behind the structure of the reviewed taxonomies is that a specific answer, or a specific set of answers, exists for a given question. Lehnert and Greaser also seem to assume that the answer is known—not necessarily by the person asking the question, in which case it would be a rhetorical question, but possibly by the person to whom the question is directed. Such questions are characteristic of *convergent* thinking, where the questioner is attempting to converge on "the facts." The answers to converging questions are expected to hold truth-value since the questioner expects the answering person to believe his/her answers to be true. Almost all of the categories of questions contained in Lehnert's taxonomy, including the ones Graesser refers to as Deep Reasoning Questions, DRQs, are converging in nature. An example is: "Why does the moon rise at night?" where the questioner is seeking a rational and truthful explanation for the rise of the moon.

However, questions that are raised in design situations tend to operate under the opposite premise: for any given question, there exists, regardless of being true or false, multiple alternative known answers as well as multiple possible unknown answers. The questioner's intention is to disclose the alternative known answers, and to generate the possible unknown ones—regardless of their truth value. Such questions are characteristic of *divergent* thinking, where the questioner is attempting to diverge away from the facts to the possibilities that can be generated from them. I find it useful to establish a terminology for these types of diverging questions, and name them "Generative Design Questions," or GDQs. An example is: "How can one reach the moon?" where the questioner wants to generate possible ways of reaching the moon, and, at the time of posing the question, is not too concerned with the truthfulness of potential answers.

A GDQ generally yields multiple answers, which satisfy the question to various degrees. Upon asking a diverging question, the designer's role is precisely to tackle that quality of it by investigating how each answer satisfies the question, and establishing criteria for favoring one answer over the others. That process of investigation, comparison, and evaluation constitutes decision making in design. And, as argued for in section 2.2.2, it does not necessarily take place after the question is posed; it also occurs *while* the question is being formulated.

Therefore, a coding scheme for analyzing the questions asked while designing needs to account for the types of questions that fall under the GDQ concept as well if it is to be comprehensive. A good starting point is to adopt one of the more established taxonomies and augment it by adding GDQ categories. Two of the taxonomies reviewed in section 2.3—Dillon's and Lehnert's—are articulate (since Graesser's taxonomy is an extension of Lehnert's, I will be referring to Lehnert only).

Although Dillon's taxonomy appears to be more structured, it is more appropriate for me to adopt Lehnert's for two reasons:

1. Lehnert's taxonomy has been proven to be effective in coding questions in discourse, and its utility as a coding scheme has been enhanced by Graesser's discussion on DRQs.
2. It might be possible and meaningful to implement aspects of the questioning framework used this study in a design information support

tool. Since Lehnert developed her taxonomy with the intention of creating an artificial intelligence that can answer questions, and implemented it as a computer program, it would be more feasible to implement a framework that is based on hers computationally.

Therefore, I used Lehnert's taxonomy of questions as the basis for the coding scheme used in this study. I then analyzed the questions that I could not account for, the GDQs, and proposed 5 new GDQ categories as extensions to Lehnert's taxonomy: Proposal/Negotiation, Scenario Creation, Ideation, Method Generation, and Enablement.

In the next section, will discuss and provide specific examples of each GDQ category. I will also illustrate the context in which each type of question occurs, and their significance, by providing transcripts extracted from data collected during the laboratory experiments.

3.3.1 Proposal/Negotiation

The questioner wants to suggest a concept, or to negotiate an existing or previously suggested concept. These types of questions initially appear to fall under the "Judgmental" category, which covers questions where the questioner wants to solicit a judgment from the answerer by requiring a projection of events rather than a strict recall of events. However, there is a fundamental conceptual difference between making a suggestion and soliciting a judgment.

An example of a Judgmental question is, "Do you think the wheel is more accurate?" The questioner is asking for the answerer's opinion on what should be done, and is not offering any opinion herself/himself. The answerer is expected to supply a single definitive opinion.

On the other hand, "How about attaching a wheel to the long LEGO piece?" is a Proposal/Negotiation question. The questioner is offering an opinion on a concept, and expecting the answerer to supply her/his own corresponding opinion(s), which would not be definitive. The questioner intends to establish a negotiation process by exchanging opinions, and to open up the possibility of new concepts. The suggestion of the new concept usually requires a consideration of the hypothetical possibilities the new concept can lead to.

Another example of a Proposal/Negotiation question is provided in Transcript 3-1, where Team 12 is considering a sensing concept for the measurement device. The consideration results in a new measurement concept.

Transcript 3-1. Design team members A, B, and C are considering a sensing concept for a measurement device. The consideration results in a new measurement concept. The Proposal/Negotiation question is highlighted in bold type.

Time	Sub	Utterance
23:49	B	What do you call that?
23:52	C	Just a roller.
23:52	A	That would be a really interesting one. Just one piece you know the diameter of.
23:54	B	Roller…
23:57	C	It's basically a roller measurement. It's the same thing they use to lay out stuff on the streets
24:05	A	Or, you can make a… (cut off by C)
24:07	C	So basically do it in turns of fractions of circumference.
24:11	B	Okay, so we have a roller and then measure how many revolutions?
24:17	**A**	**Yeah, or you can have a series of Legos connected like a linkage that's really bendable, just kind of wrap it around like a tape measure, right?**
24:26	B	That's a good idea. It's another…
24:28	C	It's kind of an end-to-end thing you're talking about? So, you basically have two lengths that pivot, you know what I'm saying? So, you kind of flip one over the other and work your way around.
24:45	A	I was just thinking like…(cut off by C)
24:47	C	I was interpreting, trying to interpret what you're saying to mean something like this where you have something like this.
24:56	A	Oh, exactly.
24:58	C	That you could work your way around and flip one over the other so that you always have on length in contact with the surface that you're trying to measure.

At the beginning of the transcript segment, C has already come up with the "roller" concept where the sensor component of the measurement device is a wheel of known diameter that rotates freely on the surface to be measured. In the next 15 seconds, A and B converse with C, and learn how the roller works. When they understand that each revolution corresponds to a known distance, A transforms the concept to a linear domain and suggests the possibility of using a series of flexible linear linkages such as a "bendable tape measure." A voices his suggestion in the form of the Proposal/Negotiation question highlighted in bold type in Transcript 3-1. C immediately responds to A's suggestion. He first makes sure he understood A's suggestion correctly, and then proceeds to refine the concept by negotiating its application method.

As can be seen in this interaction, Proposal/Negotiation questions are significant because proposing an idea in the form of a question promotes consideration and feedback, and negotiation promotes synthesis.

3.3.2 Scenario Creation

The questioner constructs a scenario involving the question concept and wants to investigate the possible outcomes. In a strict sense, such questions could be categorized under Lehnert's "Causal Consequence" category. However, Causal Consequence questions involve one causal chain of two concepts—the second concept is partially or completely unknown—joined by the LEADTO causal link. Scenario creation questions differ from causal consequence questions in two ways: there are multiple possible causal chains and linked concepts, and the causal link is CANCAUSE since the causal chains are hypothetical.

An example of a causal consequence question is "What happened when you pressed the pulley?" The questioner is assuming that when the person pressed the pulley, there was a reaction, and something specific happened. In other words, the person pressing the pulley led to a specific outcome, and the questioner wants to know what that was.

On the other hand, "What if the device was used on a child?" is a Scenario Creation question. The questioner wants to generate and account for as many possible outcomes as possible from the scenario(s) that can be constructed.

Another example of a Scenario Creation question is provided in Transcript 3-2, where Team 10 is evaluating a sensing concept for the measurement device. The evaluation results in the creation of a new measurement concept.

At the beginning of the transcript, A, B and C are evaluating a sensing concept for the measurement device, where the sensor component is a wheel of known diameter that rotates freely on the surface to be measured. A comments that the wheel rolls even on clothing. However, C realizes that it depends on how much pressure is applied on the axle of the wheel, and that it might slip. About 10 seconds later, C uses that insight to pose a Scenario Creation question, and wonders if the wheel would rotate without slipping on hair (the device will be used to measure the circumference of a human head). In essence, C constructs a new design requirement: the wheel should rotate freely and without slipping on hair. B then tests the device on his head, and reports that it indeed slips. At the end, C comes up with a new concept, which uses different size "interchangeable" wheels—the assumption being that a larger wheel would be less likely to slip.

Transcript 3-2. Design team members A, B, and C are evaluating a sensing concept for a measurement device. The evaluation results in the creation of a new concept. The Scenario Creation question is highlighted in bold type.

Time	Sub	Utterance
48:23	A	We gotta keep this from rotating.
48:30	B	Can we like bend this?
48:36	A	Oh, what is this? Hey, check this out. I wonder if this has a rolling end?
48:51	A	Even works on clothing.
48:53	C	Yeah, it really's a matter of how tight you squeeze it.
48:56	B	We can do this.
48:59	A	That cantilever is wicked though.
49:02	**C**	**What about people who have hair?**
49:04	B	(laughing) Are you making fun of my hair?
49:06	C	(seriously) No, I'm saying that we have to measure...like this little wheel wouldn't work because it's not going to roll over long hair...even on my short hair it won't work.
49:15	A	Is it rolling?
49:16	B	No, a little bit.
49:18	A	Like, it slips.
49:19	B	You can't roll my...does it... [cut off by C]
49:20	C	Whereas the big one, or we could have an interchangeable roller, one that is pop-in for head, and pop-in for the hand.
49:28	B	Yeah.

As can be seen in this interaction, Scenario Creation questions are significant because accounting for possible outcomes generates and refines design requirements.

3.3.3 Ideation

The questioner wants to generate as many concepts as possible from an instrument without trying to achieve a specific goal. Such questions involve multiple possible concepts and causal chains. The first concept is partially unknown, and the second concept is partially or completely unknown.

An example of an ideation question is, "Are magnets useful in anyway?" The questioner does not intend to achieve a specific goal by using the magnets. He/she does not have a purpose other than to generate as many ways of utilizing magnets as possible. The role of that question is illustrated in Transcript 3-3, where team 10 is considering magnets they came across while going through the hardware they were given to design and prototype the measurement device. The consideration results in a concept for holding the device while not in use.

Transcript 3-3. Design team members A, B, and C is considering some magnets they came across while going through the hardware they were given to design and prototype a measurement device. The consideration results in a concept for holding the device while not in use. The Ideation question is highlighted in bold type.

Time	Sub	Utterance
29:34	A	Wait, is this part of the kit?
29:36	B	Yes, magnets.
29:37	**A**	Hey there's magnets. **Are magnets useful in anyway?**
29:43	C	Yeah, if we wanna make an oscilloscope. (B laughs)
29:48	A	Let's try all the interesting pieces and see what we can do with them. Have an interesting piece section…I have no idea what it is…magnets…let's keep on moving them into big piles.
30:10	C	I don't even know why we have ball joints.
30:23	A	Let's see what they do here. They actually use these as rubber bands. That's kind of interesting…it would be cool to use our stuff.
30:35	C	I think these are just for these
30:38	B	What is that for?
30:39	A	Oh, that's interesting. Remember, Aesthetics count. Rubber band…(writing down the ideas)…uhm…squeeze handle, maybe we can do a squeeze handle. I don't know…Let's look through some of these cases.
30:57	B	There's something that bends.
31:07	A	Sockets just seem to stick out…Did you see the sockets do anything? They use sockets here to use the rubber bands to go on.
31:19	C	Oh.
		(all three looking through the Lego manual)
31:38	A	Looks cool.
31:41	C	Let's make it (laughs).
31:57	A	Yeah, the magnet's sitting there, but it doesn't do anything.
32:01	C	They use magnets here?
32:03	A	These are the magnets, right? With these tiny things clicked onto here. I'm not sure what they do.
32:10	B	I think it's just supposed to just hang stuff there.
32:13	C	So basically we have this thing, right?
32:15	B	Just hang stuff there.
32:18	C	That's his gun. He picks up at his pack and puts it…
32:21	**A**	So maybe we can use the magnet, maybe for as like a holder, so when you're done with it you just click it onto the wall or something…**What else can we do with magnets?**

At the beginning of the transcript segment, A identifies the magnets, and immediately poses an Ideation question in order to generate concepts for using them. It is important to note that at that point, A is acting without a specific goal; he does not have a specific role for magnets in mind. For a few seconds, they get distracted and focus on other "interesting" pieces like magnets such as ball joints, rubber bands, and sockets, but they quickly return to the magnets and examine how they are used in the LEGO kit the parts came from. What they learn influences A to consider magnets as a part of a concept for holding the device while not in use. As soon as he generates this concept, he poses the same Ideation question to generate more concepts.

As can be seen in this interaction, Ideation questions are significant because operating without a specific goal frees associations and drives concept generation.

3.3.4 Method Generation

The questioner wants to generate as many ways as possible of achieving a specific goal. Even though such questions initially seem to be derivatives of Lehnert's "Procedural" category, they are conceptually different. As Lehnert points out, "A Procedural questions asks about an act that was simultaneous with the main act of the question. If a question asks about an act that precedes the main act of the question, the question is either a Causal Antecedent or an Enablement question." A method generation questions falls into the second category since it asks about acts that precede the main act of the question. Then, according to Lehnert, it should be classified as a Causal Antecedent or an Enablement question. However, Causal Antecedent and Enablement questions each involve a single causal link, whereas a method generation question has a completely known initial question concept and multiple possible and completely unknown secondary question concepts.

An example of a method generation question is, "How can we keep the wheel from slipping?" The questioner wants to generate secondary concepts, which, if realized, will cause the initial concept—keep the wheel from slipping. That question is clearly distinct from the causal consequence question, "What happened after you pressed the pulley?"

Another example of a Method Generation question is provided in Transcript 3-4, where Team 5 is generating methods for implementing an automatic readout of the measurement device. The evaluation results in the creation of several new readout methods.

At the beginning of the transcript segment, A invites the team to brainstorm readout methods. He immediately poses a Method Generation question, and sets their goal, which is to generate new methods for implementing an automatic readout, where the measurement the device takes is indicated in such a way that all the user needs to do is to look at the readout and read it off. The team responds, and within 60 seconds, generates 3 different methods.

Transcript 3-4. Design team members A, B, and C are generating methods for implementing an automatic readout of a measurement device. The evaluation results in the creation of 3 new readout methods. The Method Generation is highlighted in bold type.

Time	Sub	Utterance
05:01	A	Let's brainstorm read-out methods. New topic. **However you measure it, how can you make it automatically readable?**
05:16	B	Okay, so have the audible clicking.
05:19	C	I think if we can do a visual.
05:22	A	Is there a rack and pinion? No, just simple gears.
05:28	C	We have some bevel gears though. I don't know if it's...
05:32	A	But if the spur gear rolls along a page, you can then whip out a tape measure and say, okay, this is how far it went, or something like that. You can make it like roll along something else.
05:44	B	That's why I was thinking if we wound up the string when you made the measurement then you just unroll the string and measure it...The rod I think is better. That's not elegant—unwinding some string and measuring it.
06:08	B	There might be way to make a magnet flip like 180 degrees every time.

As can be seen in this interaction, Method Generation questions are significant because operating with a specific goal generates a set of methods for implementing concepts.

3.3.5 Enablement

The questioner wants to construct acts, states, or resources that can enable the question concepts. This category is the GDQ version of the original Enablement category Lehnert proposed, which Graesser labeled as a DRQ. What differentiates it from Lehnert's, and makes it a GDQ, is the questioner's assumption of multiple possible initial concepts.

An example of a GDQ Enablement question is, "What allows you to measure distance?" if the questioner is indeed aiming at identifying resources for measuring distance. However, the same questions should be categorized as a DRQ enablement question if the questioner believes there is a single or a set of specific known resources of measuring distance. That differentiation can only be made by taking into account the context in which the question was posed.

Another example of an Enablement question is provided in Transcript 3-5, where Team 7 is generating resources that enable the implementation of a measurement concept. The evaluation results in the identification of an existing resource and the generation of a new one.

At the beginning of the transcript, B poses an Enablement question in order to generate resources that can rotate and measure distance. It is important to note that he already has a measurement method in mind—

rotation—and that he is looking for enabling resources. B immediately answers his own question by identifying a tape measure as a possible resource. Influenced by the tape measure idea, A then considers a different measurement method—conforming a series of linkages to the measurement surface—and generates a new resource that would enable it, consisting of a straight LEGO pieces of known length connected at the ends. B briefly considers A's idea, and then returns to the Enablement question he asked to generate more resources.

Transcript 3-5. Design team members A, B, and C are generating resources that enable a measurement concept. The evaluation results in the identification of an existing resource and in the generation of a new one. The Enablement question is highlighted in bold type.

Time	Sub	Utterance
21:05	**B**	So, **what goes around a circle and measures things?** You know...when you...like you ever...(pause)...Tape measure's pretty good. A tape measure!
21:20	C	I just keep thinking you just rotate this thing around.
21:25	A	Not necessarily. We can have something like let's say if we have a lot of little pieces joined like this, right...we can actually just put it around the hand. And it won't be...we'll have some minor error because it has spaces here, but if we do that we want each one like let's say this is one inch...these are all certain inches, certain lengths...we can just put that around the hand and measure how long it is.
21:51	B	I guess my comment, like things, my concern is that's a lot of parts, be we shouldn't really, we shouldn't really limit ourselves right now. But let's see what else we have.

As can be seen in this interaction, Enablement questions are significant because identification of multiple resources promotes surveying and learning from existing design features.

3.4 Comparison of the Taxonomic Approaches

There are striking similarities between the taxonomies reviewed in section 2.3. I already mentioned that Kuffner's and Baya's frameworks are rather similar. That is mainly because they both adopted highly focused and similar information-centric views. However, as Graesser argued when mapping Lehnert's taxonomy of questions to Bloom's taxonomy of educational goals, information-seeking questions have a lower significance in learning than the more sophisticated analysis and synthesis questions. It can be argued this is the case for designers as well. Therefore, understanding more about design thinking requires the construction of a taxonomy of questions that goes beyond accounting for information-seeking questions.

At this point, it is appropriate to revisit and compare the classification schemes of Aristotle, Dillon, and Graesser, and my extensions to Lehnert's scheme. In section 2.3, I discussed Aristotle's influence on Dillon's approach, and the mapping between their schemes. I also discussed the origins of Lehnert's approach, and its adoption and extension by Graesser through the addition of five new categories. I remarked that Greaser's identified a class of questions as Deep Reasoning Questions (DRQs), which are related with learning performance.

In section 3.3, I discussed the rationale for basing my coding scheme on Lehnert's (including Graesser's extensions), and argued that five more additional categories representing divergent thinking—Generative Design Questions (GDQs)—were necessary for it to be applicable to design situations. Thus, what we have so far is two parallel evolutionary threads on the taxonomy of questions. What remains is to compare them to see if they map onto each other.

The comparison can be conducted by inserting the five taxonomies into the columns of a table, and attempting to align the rows—the categories—that are similar in nature. Mutually populated rows would indicate synergy between the schemes. Table 3-1 illustrates the result of that comparison.

As Dillon pointed out, the differentiating principle between his and Aristotle's question categories is the extent of "knowledge about some phenomenon P entailed in answer." The hierarchy is the natural progression of that knowledge; the lower categories of questions contained in the initial classes have less knowledge in their answers than the higher categories of questions contained in the latter categories. The categories of questions contained in the last class have no, or unspecified, knowledge in their answers that is directly provided by the answerer (with the exception of the Deliberation category). Therefore, their positioning is irrelevant. Before discussing the appropriate positioning of the fifth class of questions, I will focus on the first four and the sixth classes, and determine if the schemes map with respect to them.

Looking at Table 3-1, it is immediately apparent that Lehnert's scheme is missing the Instance category under the Existence class, the entire Nature class, the Equivalence and Difference categories under the Fact class, and the Relation and Correlation categories under the Reason class. On the other hand, Dillon's scheme does not articulate the Procedural/Instrumental, Enablement, and Judgmental categories that Lehnert's scheme contains. The rest of the categories in Dillon's and Lehnert's schemes map well.

The unaddressed Nature class in Lehnert's scheme is addressed in Graesser's by the Definition and Example categories, and the Equivalence and Difference categories under the Existence class by the slightly broader

Comparison category. Although Graesser's scheme does not directly address the Relation and Correlation categories, it can be argued that his Interpretation category partially maps onto them; interpretation questions can be thought to be exploring relationships and correlation between phenomena in order to construct causal explanations and projections. Also, the Enablement and Procedural/Instrumental categories not articulated by Dillon's scheme are most likely implied in Aristotle's Reason class, since such questions must assume and operate on the basis of causality.

Table 3-1. A visual comparison of the categories of five taxonomies of questions. Dillon's categories are an expansion of Aristotle's. Graesser's and Eris's categories are an extension of Lehnert's. The types of questions termed "Deep Reasoning Questions" by Graesser are italicized. The types of questions termed "Generative Design Questions" by Eris are in bold.

ARISTOTLE	DILLON	LEHNERT	GRAESSER	ERIS
Existence (Affirmation)	Existence	Verification	Verification	Verification
	Instance			
Nature (Essence)	Substance		Definition	Definition
			Example	Example
Fact (Attribute/ Description)	Character/ Description	Feature Spec.	Feature Spec.	Feature Spec.
		Concept Complete	Concept Complete	Concept Complete
		Quantification	Quantification	Quantification
	Function	Goal Orientation	*Goal Orientation*	*Rationale/Function*
	Rationale			
	Concomitance	Disjunctive	Disjunctive	Disjunctive
	Equivalence		Comparison	Comparison
	Difference			
Reason (Cause/ Explanation)	Relation		Interpretation	*Interpretation*
	Correlation			
	Conditionality & Causality	Causal Antecedent	*Causal Antecedent*	*Causal Antecedent*
		Causal Consequent	*Causal Consequent*	*Causal Consequent*
		Expectational	*Expectational*	*Expectational*
		Procedural	*Procedural*	*Procedural*
		Enablement	*Enablement*	*Enablement*
				Proposal/Negotiation
				Enablement
				Method Generation
				Scenario Creation
				Ideation
		Judgmental	Judgmental	Judgmental
	Rhetorical		Assertion	
		Request	Request/Directive	Request
	Deliberation			

Dillon's scheme does not address the Judgmental category proposed in Lehnert's scheme. That is mostly likely the result of Dillon's focus on research questions. When considered within the scope of Lehnert's framework, the Judgmental category is difficult to position among the other categories; all questions are judgmental questions to some extent since a question cannot be answered based purely on "fact" or with complete "objectivity." Therefore, I decided to treat the Judgmental category as a specific class, and to position it below the first five classes that are conceptually related.

In conclusion, at a fundamental conceptual level, the version of Lehnert's scheme Graesser augmented maps onto Dillon's, and thus, onto Aristotle's scheme. That is a positive finding as it indicates a strong degree of agreement in the thinking of the authors, and assures me that Lehnert's framework constitutes a sound basis for my analysis.

The fifth class of questions in Table 3-1 containing the Generative Design Questions is the contribution of this research. It is not addressed by any of the other schemes. For the most part, this can be explained by the diverging-converging thinking paradigm I argued for in the previous section, where I made a fundamental distinction between questions that aim to converge on facts, and questions that aim to diverge away from facts to the possibilities that can be generated from them. The classification schemes of Aristotle, Dillon, Lehnert and Graesser are concerned mainly with convergent questions.

One way of supporting that claim is to analyze each question category according to the convergent-divergent paradigm. A more abstract, yet equally valid, way of supporting the claim is to consider the motivations of the authors for constructing the taxonomies, and to determine if they aim to establish frameworks for understanding facts, or for creating possibilities from facts. Aristotle's paradigm is epistemological; as I remarked earlier, his main premise was: "The kinds of question we ask are as many as the kinds of things which we know." Thus, he focused on what we know, on the existing, and not on the possible. Dillon explicitly stated that his taxonomy is descriptive of "research" questions, and his interpretation of research activity seems to entail discovery and better understanding of naturally occurring phenomena, paralleling Aristotle's paradigm.

And finally, Lehnert, strongly influenced by cognitive science, was ultimately interested in developing a question answering process, consisting of two separate processes for understanding questions and finding answers. The second process of "finding"—not creating—answers entails retrieving answers from existing memory structures. (Even though she mentions that multiple appropriate answers can be constructed for most questions using that procedure, that should not be taken to mean that possibilities can be created from known facts; it means that multiple known answers might exist and can be "found" in the memory structure.)

On the other hand, as I argued for in the previous section, the Generative Design Question categories I propose reflect divergent thinking. I therefore form a separate class of questions from them. However, it is not necessarily clear where that class should be positioned in Table 3-1 because hierarchy expressed in the table is determined by the extent of knowledge in the answers.

Does the knowledge in answers of GDQs encompass the knowledge in answers of the other class of questions? That is a problematic proposition since the purpose of GDQs is to create knowledge as opposed to discover or to construct knowledge based on fact, and it is inappropriate to guess at the extent of knowledge that is yet to be created before it is created. At this point, I can only hypothesize that GDQs, similar to DRQs, are correlated with learning, and also that both GDQs and DRQs are correlated with design cognition, and, thus, with design performance. Verifying that hypothesis would imply that the extent of knowledge in answers to GDQs is comparable to the extent of knowledge in answers to DRQs, and to the types of questions in Aristotle's Reason class. I will address this hypothesis throughout the empirical dimension of this research in the following chapters.

Chapter 4

HYPOTHESIS GENERATION IN THE FIELD: SHADOWING THE DESIGN TEAM

The empirical dimension of this research consists of three progressive steps:

1. Observation and analysis of a realistic design project in the field for hypothesis generation.
2. Design of a laboratory experiment to test the hypotheses.
3. Redesign of the experiment and the execution of the final version.

This empirical design research approach—segmenting the research project into three progressive steps—has been practiced at the Stanford University Center for Design Research since the late 1980s. It identifies a conceptual progression by structuring the empirical dimension of a research project into three sequential research components that build on each other, and by characterizing the scope and outcome of each component.

In order to provide more structure for each of the three steps, I relied on another approach that has been effectively used at the Center for Design Research[7]. It entails the iteration of a cycle consisting of the "Observe-Analyze-Intervene" phases, and advocates going beyond merely observing and describing design activity to constructing meaningful interventions that test gained insights (Figure 4-1).

---

[7] This method is too generic to be attributed to an individual. However, at the Center for Design Research, it was first used by Tang and Minneman [Tang 1989, Minneman 1991].

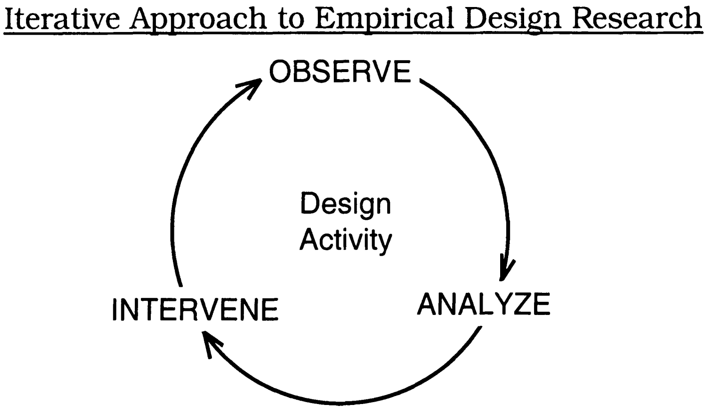


Figure 4-1. The iterative approach to empirical design research entails a cycle consisting of the "Observe-Analyze-Intervene" phases, and advocates going beyond merely observing and describing design activity to constructing meaningful interventions that test gained insights.

In order to use the two approaches in conjunction, I superimposed the iterative approach on each of the three empirical steps. Within each step, I conducted multiple iterations of the cycle. The differences in the nature of the empirical steps require more or less emphasis on the different phases of the cycle [Figure 4-2].

Specifically, during hypothesis generation, it is not useful—even counterproductive—to focus on intervention. The main purpose is to observe and understand the design situation and the phenomena of interest in the field. The goal of designing a laboratory experiment is to incorporate the understanding gained during hypothesis generation into experimental elements such as a design scenario, research variables, and a meaningful intervention, and create a pilot experiment. The final empirical step involves running the pilot experiment, observing and analyzing the experimental elements, and redesigning them to achieve the intended intervention. The redesigned experiment is then conducted and the data are analyzed in depth.

In this chapter, I address the first step of the empirical dimension of this research, hypothesis generation in the field. The other steps are addressed in Chapters 5, 6, and 7. In section 4.1, I discuss the grounded principles used in hypothesis generation. In section 4.2, I provide the context for the preliminary field observations. In section 4.3, I outline and compare two techniques for capturing design activity in the field. In section 4.4, I report the findings of the field research, which include key observations and a set of hypotheses.

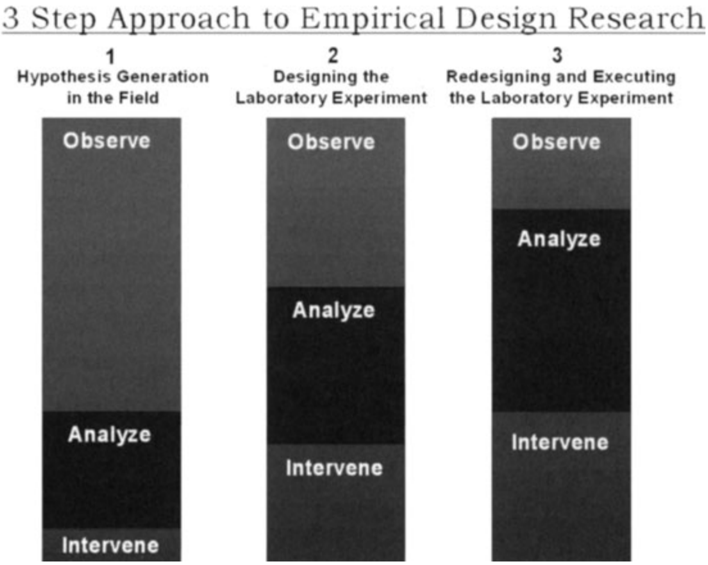


Figure 4-2. The "Observe-Analyze-Intervene" cycle superimposed on the three steps of the empirical dimension of the research. Each step entails multiple iterations of the cycle. Differences in the nature of the empirical steps require more or less emphasis on different phases of the cycle. The relative dimensions of the bars for each step are approximations for the time spent during each phase.

4.1 Grounded Principle for Hypotheses Generation

In order to generate hypotheses in the field, I used a grounded approach, which involves identifying a realistic design situation, and observing and capturing the activity in various forms for analysis.

The grounded approach bases the observations in design practice, and ensures that the resulting hypotheses are relevant. If the researcher brings his/her viewpoint into the process too early, the resulting hypotheses run the risk of being unsound and irrelevant. And, naturally, verifying irrelevant hypothesis through experimentation accomplishes little in advancing our understanding of design activity.

In other words, it is absolutely necessary to study design activity first—regardless of one's prior knowledge of the phenomena under observation. Although this principle sounds rudimentary, it is easy for design researchers to inadvertently drift away from it while observing "others" design, and develop a position on what "should be done."

I believe there are two reasons why this tends to happen:

1. Unlike social scientists studying social phenomena, design researchers studying design activity—a socio-technical phenomenon—tend to be practicing designers, and, in many cases, engineers. And, unlike social scientists, designers and engineers are trained to intervene, change, and create systems rather than to solely observe and understand them. (That is not to say designers and engineers are not trained to observe and understand, but to say that the context of their observations, and hence, their primary intent, is to intervene and create change.)
2. On a more speculative note, the nature of the activity under observation, designing, is simply engaging. If one were to observe swimmers swim, one would not necessarily be tempted to start swimming himself/herself. However, if one is observing designers design, the sensation is rather different as design activity has an encompassing human quality that invokes participation.

Therefore, applying grounded principles to empirical design research can require the researcher to constantly remind himself/herself of such influences while observing design situations.

4.2 Context of the Preliminary Observations

It is necessary to provide some context for the preliminary observations I made during hypotheses generation (the observations are presented in section 4.3). Therefore, in this section, I will briefly discuss the setting for the observations, the designers I observed, and the design project they were working on.

4.2.1 The Setting: Mechanical Engineering 210, a Graduate Level Design Class

The setting for the preliminary observations was a graduate level engineering design class at Stanford University, Mechanical Engineering 210, Mechatronics Systems Design[8]. The class lasts an academic year (three academic quarters), and typically involves 30-40 students working in teams of 3-4 on industry sponsored design projects. Students are exposed to and master state of the art design processes and design support technology. In

---

[8] The observations of ME210 provided in this chapter are based on the version offered in the 1998-1999 academic year.

order to accelerate learning, a socio-technical infrastructure consisting of extensive coaching resources and collaborative design tools, is deployed. A "design loft," a communal workspace, where each team has a designated open work area, facilitates interaction and integration of resources.

During the first quarter of the class, students go through numerous warm-up design exercises in teams. At the end of the second month, they are introduced to a pool of industry sponsored projects, finalize their team formation efforts, and choose a project. Each industry sponsor provides conceptual and logistical assistance via a project liaison, and financial assistance in the form of a $15,000 budget per team. At the end of nine months, the teams are expected to deliver a functional prototype as well as detailed documentation of the design they have developed. The class has a history of producing highly successful projects (as measured by the success rate at the national Lincoln Arc Welding design competition).

Apart from its educational value to students, this setting has also served as an observational platform and a test bed for researchers at the Center for Design Research. Since the class is structured to simulate a realistic design environment—resembling an industrial setting—the design activity it promotes can be treated as valuable and relevant data [Mabogunje 1997]. It can also serve as an experimental space where innovative design support tools can be introduced and tested[9].

4.2.2 The People: A Four Person Design Team

The ME 210 design team I observed was made up of four graduate mechanical engineering students with mechanical engineering backgrounds. They were taking ME 210 as their core design class in the masters program. The team composition was in accordance with the design team-construction method developed by Wilde, which takes academic and psychological descriptors of team members into account in forming an academically and socially balanced team [Wilde 1997]. The team I studied was unusual in one aspect: it consisted of three females and one male (in a field where male to female ration is often above 10-1). The team members did not know each other before attending the class, and formed their team using Wilde's team formation guidelines approximately two days before I began to observe them.

---

[9] There are ethical issues associated with this approach that require careful consideration.

4.2.3 The Project: Design, Build and Race a Paper Bicycle

Prior to the introduction of the industry sponsored projects, students participate in a two week long introductory design exercise, which serves as a warm-up and orients the students with the methodologies and technology that will be used throughout the year. For more than five years, the design task used in the introductory exercise had been to design, prototype and race a paper bicycle. The final prototype is expected to be built mainly out of paper components, and meet weight, durability, and stability constraints. At the end of the two weeks, the teams enter a bicycle race with their prototypes, which takes place around a 400 feet circular track. Even though the duration of the exercise is somewhat short, it is still a valid source of preliminary data for hypothesis generation.

4.3 Two Techniques for Capturing Design Activity in the Field and Generating Hypothesis

I relied on two techniques while gathering data in the field and generating hypotheses. Since both techniques are well established, I will not describe them in detail. Instead, I will consider their use in empirical design research.

4.3.1 Ethnographic Approach: Shadowing the Design Team

In Designing Engineers, Bucciarelli used ethnographic techniques in developing a social theory of design, and discussed their use in observing engineering design situations [Bucciarelli 1988, 1994]. He pointed out that ethnographic techniques are an effective way to move beyond understanding designing simply by studying products to understanding designing by studying the design activity the products are created in. Therefore, ethnography is an effective methodology for abiding by the grounded principle outlined earlier.

Before utilizing ethnographic techniques in the field, it is imperative to ensure the feasibility of observing the design situation one wants to study. For instance, innovative commercial design projects are typically under tight confidentiality regulations, and access to the "activity" is permitted only in certain conditions. It is important to consider the effects these limitations might have on the study as some situations simply do not permit the level of access necessary to generate significant insights. In most cases, such limitations can be negotiated and reduced over time.

Fortunately, the setting for the field observations in this research, ME 210, did not pose any significant limitations as graduate students tend to be open to observation. However, even though the course strives to "simulate" realistic design situations, what takes place in the class still possesses an academic quality. It is possible to view this as a tradeoff between access and reality. In an academic setting, the researcher has nearly unlimited access, but less real-life data. The converse is true in an industry setting.

For the purposes of this study I, together with a colleague, "shadowed" a four person ME 210 team during the paper bicycle project. Upon spending a brief amount of time with each team prior to the beginning of the project, we chose a team we thought would be the most accessible. The team agreed to inform us in advance of the time and place of their informal and formal group meetings—design sessions. Over the two week duration of the project, we were notified of over nine design sessions, and observed all of them with ethnographic techniques.

4.3.2 Video Interaction Analysis: Generating the Hypotheses

Another technique we employed in conjunction with ethnography was to capture the interaction during the design sessions with a video recorder. Fundamentals of video interaction analysis and its use in design research have been discussed by Tang and Cross [Tang 1991, Cross 1996].

A significant difference between ethnography and video interaction analysis is that, as an ethnographer, the researcher relies on his own senses and strives to document as much of his perceptions as possible during and after the observations, whereas when using the video camera, the researcher relies on the audio and video information the video camera can capture. Therefore, each method serves to document the activity through a different "lens." This is desirable since, if used in conjunction, the data generated by each technique can be complementary—the findings generated with one method can add clarity and meaning to the findings generated with the other.

Another significant difference between the two techniques is that the information captured with a video recorder can be replayed. This has two implications: video data can be shared and independently analyzed by other researchers who did not directly observe the captured design activity; and when aiming to generate hypotheses, video data can be jointly analyzed by a group of researchers to facilitate unstructured reflection.

The first implication widens the scope of data analysis that can be conducted. As was the case with the data the book Analyzing Design Activity was based on [Cross 1996], videotapes can be sent to groups of researchers for analysis and interpretation. The findings can then be compared and synthesized into a collective understanding with broader implications.

In the second implication, what I mean by "unstructured reflection" is a form of collective brainstorming. Several researchers watch the videotapes together, and, while doing so, speculate freely on any aspect of the activity that might attract their attention with the intent of generating hypotheses. This process widens the range of interesting phenomena that can be identified as the interaction between researchers is very likely to stimulate their ideation process. This is how the audiovisual data collected during the paper bicycle project were analyzed.

4.4 Findings of the Field Research

The findings of the first empirical step are discussed in the next four sections. In the first section, I evaluate the effectiveness of the two observation and analysis techniques discussed in section 4.3. In the second section, I focus on the outcomes of the observation and analysis, and highlight four key observations. In the third section, I derive three testable hypotheses by considering the key observations together with the conceptual framework I developed on the nature of questions in Chapter 3. And finally, in the forth section, I synthesize the phenomena outlined in the hypotheses into an analytical framework for understanding and measuring design performance.

4.4.1 On Capturing Design Activity in the Field

The two techniques discussed in section 4.3 proved to be highly effective in capturing design activity in the field. Although I cannot comment on their individual effectiveness, using them in conjunction with each other enhanced the accuracy and depth of my observations by providing different levels of granularity and focus. I will illustrate this point by highlighting two common situations that a design researcher may be faced with when analyzing this type of data.

Several *tacit* elements of the interaction, which were not necessarily reflected in the videotapes, were visible when observing the activity in person. For instance, it was possible to gain a sense of the shared perspective and "mood" of the team by watching the videotape of a meeting. However, it was difficult to identify how they had evolved into their recognizable state. On the other hand, witnessing the interaction in person enabled me to sense and understand more about the perspectives and sentiments of the individual designers, and how they led to collective phenomena. What I refer to as the perspective and mood of the team were manifested strongly in the motivations, questions, choices, and the overall design thinking of the team, and, therefore, were highly relevant to the study.

Another tacit element of the interaction was what took place outside of the frame of the camera. The environment and activity in the background influenced the actions of the team. For instance, the team members often looked and pointed at artifacts—usually paper bicycles that had been designed in the preceding years—on the other side of the design loft, and discussed them. Also, there were stretches of time where one or more of them moved away from the others, and could not be captured with the video camera. What they were doing while they were away from the others, and the significance of those actions could only be interpreted by being there.

Conversely, observing interactions that were subtle, or happening simultaneously with other interactions, in person proved to be difficult since, as an ethnographer, it was only possible to focus and observe a limited number of actions at any given time. However, the video camera does not have the same limitation as an instrument; every interaction visible within its plane of focus is recorded at the same resolution, and the interaction that has been recorded can be replayed and studied for an unlimited number of times.

Therefore, while analyzing videotapes, I was able to notice interactions that I had not noticed when observing in person. For instance, it was possible to miss what a team member was doing with the prototype from a previous project while trying to follow what another one was sketching on the board. It was only when I viewed the videotape later that I noticed the interaction which had taken place between the team member and the prototype. Also, in many instances when team members were talking simultaneously within the team, or having separate one-on-one discussions, it was impossible to follow all of what was being said. Analyzing such situations from videotape enabled me to identify significant ideas, questions, and decisions that had been discussed which I had missed as an ethnographer.

What I have reported above indicates that design activity is inherently rich and can be observed and characterized at various levels. The spectrum of activity and environment depicted in Figures 4-3, 4-4, and 4-5 reflect only a fraction of that richness. The figures contain frozen "frames" from sections of the video data corresponding to progressive phases of the paper bicycle design project that I observed.

Figure 4-3. Frames from video data: The paper bicycle design team conceptualizing in their team space (on the left) and the communal design space (on the right) in the Design Loft.

Figure 4-4. Frames from video data: The paper bicycle design team exchanging ideas and best practices with another team (on the left). The paper bicycle design team prototyping their design in the Design Loft (on the right).

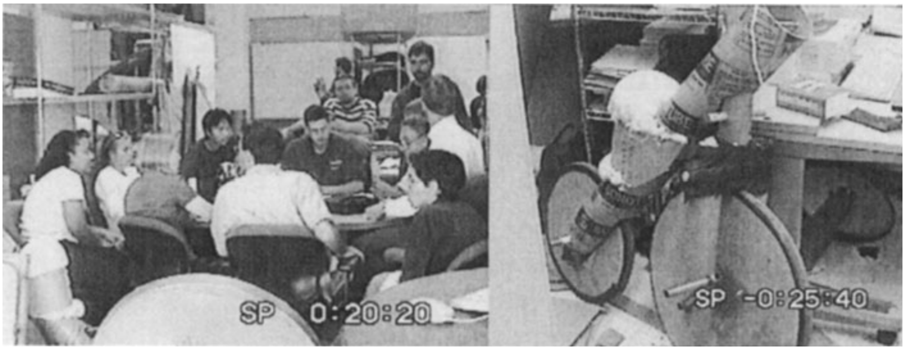


Figure 4-5. Frames from video data: The paper bicycle design team during the final design review with class TA's and instructor (on the left). The final paper bicycle prototype of the design team (on the right).

The point I would like to make with the figures is that "shadowing" the design team in the field for the duration of the project, utilizing both ethnographic and audiovisual recording techniques, and analyzing my field notes in conjunction with the videotapes allowed me to deal with the "totality" of the design activity, and gain a fundamental understanding of "what took place."

4.4.2 Key Observations

When analyzing my field notes and the videotapes, I focused on the questions that were asked, and how they influenced the interaction of the design team. What I observed was instrumental in shaping the initial concepts behind this research, and seeded many of the arguments I present throughout this book.

I made four key observations in the field:

O1: The design team members spent a significant portion of their time asking and discussing questions related to the design tasks at hand. They used questions in order to: mediate their social interaction, verify and clarify facts and each others views, seek new information, reason about and explain phenomena, and generate new concepts. (This observation alone convinced me that question asking was a subject that should be studied.)

O2: Meetings during which the team seemed to ask more "good" questions yielded more progress in terms of the insights the team seemed to gain and the discoveries they made. (At that point, in my mind, the definition of a good question was highly intuitive and subjective. It will be discussed in depth in Chapter 6.)

O3: Working with existing artifacts and prototyping hardware seemed to have an effect on the types of questions that were asked. Initially, when hardware was not present or rarely referenced, the questions were more conceptual and abstract, required long answers, and led to detailed discussions. Toward the end of the project, when the team members were discussing existing artifacts and working with prototyping hardware, the questions were much more specific and focused. (I was able to witness this trend since we had videotaped all of the meetings for the complete duration of the project.)

O4: However, identifying questions in discourse was difficult, and at times, rather problematic. I repeatedly found myself rewinding the tape after viewing the activity that followed a question just to make sure what I initially thought was a question was indeed a question.

4.4.3 Three Testable Hypothesis

The observations outlined in the previous section and the conceptual understanding I gained while developing a taxonomy of questions applicable to design activity formed a basis for generating testable hypotheses.

A good starting point was to identify elements of question asking that could be characterized and formulized. I postulated that the following two elements can be characterized in a meaningful way: the nature and the timing of a question. When I considered those conjectures in light of the first observation, O1, I wondered if they could be treated as descriptive characteristics of the design process. In other words, can a person who is exposed to these two characteristic elements of questions that are asked in a design meeting, and the content of those questions, reconstruct the fundamentals of how the team structured its design tasks? This constitutes my first hypothesis.

When I considered the second observation, my focus shifted to possible relationships between the incidence of questions and design performance. Do design teams which question more perform better? And if so, can questioning be treated as a real-time design team performance metric? This constitutes my second hypothesis.

This hypothesis is of particular importance; although many researchers agree that real-time design performance metrics are needed, none have been identified yet. There are various performance metrics which evaluate products of design activity such as sketches, documentation, system specifications, and designed artifacts. However, when compared to a real-time metric, product-based metrics are of lesser utility in understanding and managing an *ongoing* design project.

The third observation led me to consider the potential effects of working with prototyping hardware on the question asking behavior of designers. I assumed that the observed changes in the nature of the questions asked would be reflected in their "type" if they were to be categorized according to the framework developed in Chapter 3. By integrating that assumption with the third observation, I postulated that the types of questions design teams ask change when they transition from working in the absence of hardware to working with hardware. This constitutes my third hypothesis.

To summarize, O1, O2, and O3 led to the following testable hypotheses:

H1: Question timing and type are descriptive characteristics of design cognition and process. When the set of questions a design team asks during a design project is considered as a whole, the timing and nature of those questions point at the fundamentals of the knowledge and rationale the team uses for breaking down and structuring the project into design phases. Question timing and type are informative enough to serve as a roadmap to the design thinking and process of the team.

H2: Overall question asking rate is related to design team performance and can be taken as a design performance metric. There is a strong correlation between the frequency of questions and design team performance.

H3: Question asking behavior of design teams is influenced by their access to hardware. The types of questions design teams ask change when they transition from working in the absence of hardware to working with hardware.

4.4.4 A Framework for Measuring Design Performance

When viewed together, the phenomena outlined in the hypotheses form the hierarchical elements of an analytical framework for understanding and measuring design performance (Figure 4-6). Each phenomenon can be viewed as a descriptor of a higher encompassing phenomenon. The feasibility and accuracy of treating a descriptor as a performance metric increases with decreasing level of abstraction because lower level descriptors possess more detail, and are easier to identify and measure.

It is important to note that I consider design process and design cognition to be descriptors of the same level. They are strongly dependent on each other in the sense that they feed into each other in a cyclic fashion; design cognition and process are inseparable. Individual designers, design teams, and, as I have argued for in an earlier article [Eris 2002], product development organizations, extract and construct new design processes from existing design knowledge and thinking, and the resulting design processes form the basis of new design knowledge and thinking.

The implication is that, in the context of measuring design performance, observing and testing the relationship between one of them and question asking can be considered to be sufficient in generating indirect evidence for the relationship between the other and question asking. However, in general, design processes of teams and organizations are much more transparent, and, thus, easier to observe and track than their design cognition. Therefore, when dealing with H3 during data analysis, I focused on and observed the design processes of the teams only.

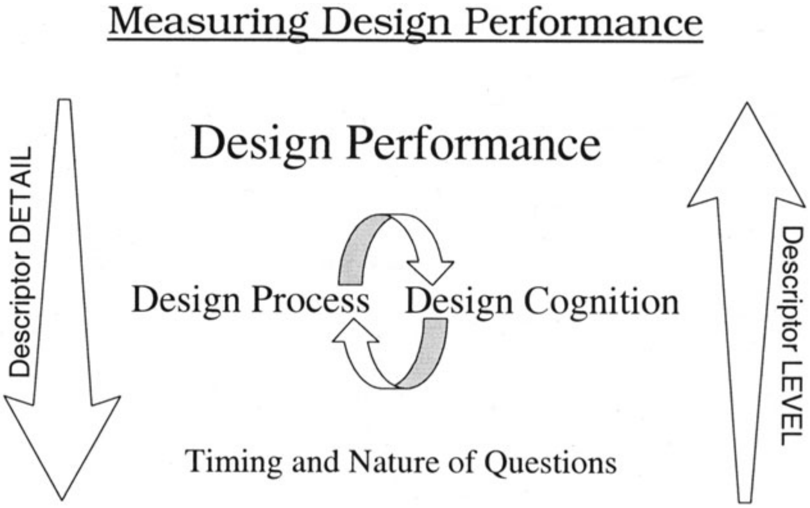


Figure 4-6. When viewed together, the phenomena outlined in the hypotheses form the hierarchical elements of a framework for understanding and measuring design performance. Validation of the hypotheses would imply the validation of this framework.

Finally, since the elements of the framework I propose for understanding and measuring design performance are hierarchical, validation of the hypotheses would imply validation of the framework as well.

Chapter 5

DESIGNING THE INTERVENTION: DIFFERENTIATING DESIGNING FROM PROBLEM SOLVING

The second empirical step of this research is designing a laboratory experiment to test the hypotheses generated during the analysis of the field observations. In the first section of this chapter, I identify and discuss seven design requirements, which can be placed under three criteria that need to be satisfied for the experiment to test the hypotheses. These criteria are: the hypotheses outlined in Chapter 4, the taxonomy of questions developed in Chapter 3, and experimental considerations specific to design research discussed in this chapter. In the second section, I discuss and propose ways of meeting each of the requirements. In the final section, I specify a design exercise that satisfies all of the requirements.

It is important to note that the analysis of the requirements under the third criterion is driven by the position that designing is distinct from problem solving, and that the experiment needs to promote the former if it is to simulate a realistic design situation. Characterizing and addressing this distinction has implications not only for this study, but also for design research experimentation in general.

5.1 Deriving Requirements for the Design Experiment

The most effective way of specifying an appropriate design experiment for testing a set of hypotheses is to *design* it. That entails identifying design criteria, expanding on those criteria by formulating design requirements, addressing each requirement individually, and integrating the resulting understanding into a unified set of specifications for the experiment.

In this research, there are seven requirements that can be placed under three experimental design criteria:

Taxonomy Related Requirement

R1: The design experiment should promote realistic question asking behavior so that the application of the taxonomy of questions, which itself is derived from data on realistic question asking behavior, is meaningful.

Hypotheses Related Requirements

R2: Definitions and metrics for the phenomena outlined in the hypotheses should be developed prior to the execution of the design experiment.

R3: The design experiment should incorporate an intervention that results in a clear distinction between design teams working with and without hardware.

Design Research Experimentation Related Requirements

R4: The design experiment should promote designing as opposed to problem solving.

R5: The setting and scenario of the design experiment should allow for the insertion of control elements associated with the hypotheses without overconstraining the activity (quasi-control as opposed to tight control).

R6: The design experiment should facilitate the testing of all hypotheses in a single experiment.

R7: The data collection methods used in the design experiment should result in data that can be analyzed qualitatively as well as quantitatively.

In each of the following three sections, I will focus on a criterion and present the rationale behind the requirements that are associated with it.

5.1.1 Taxonomy Related Requirement

R1: The design experiment should promote realistic question asking behavior so that the application of the taxonomy of questions, which itself is derived from data on realistic question asking behavior, is meaningful.

R1 reflects the understanding I gained while developing the taxonomy of questions. If the question asking behavior of the teams in the experiment are indeed realistic, and if H1 is true, then it should be possible to identify and differentiate the questions asked by the teams in terms of the categories of the taxonomy; the distinctions embodied in the taxonomy should serve as a comprehensive coding scheme for data analysis.

In other words, if the taxonomy developed in Chapter 3 is indeed comprehensive, when applied to a design situation simulating realistic design activity, each of its categories, serving as analysis codes, should receive multiple hits. And conversely, if the experimental situation indeed simulates realistic design activity, when coded by the categories of a comprehensive taxonomy, it should incur multiple hits on each category. However, the coding scheme eliciting multiple hits per category does not necessarily mean that the design situation simulates realistic design activity or that the taxonomy is comprehensive. That can only be ensured through qualitative assessment.

5.1.2 Hypotheses Related Requirements

R2: Definitions and metrics for the phenomena outlined in the hypotheses should be developed prior to the execution of the design experiment.

R2 necessitates the development of working definitions and metrics for the phenomena outlined in H1, H2, and H3 prior to conducting the experiment. Since the phenomena constitute analysis dimensions, it is important that they are characterized clearly in order to ensure that a sound data analysis framework is established before data collection takes place. The phenomena under investigation are:

1. Question Timing and Frequency
2. Question Type
3. Design Phase
4. Design Team Performance

R3: The design experiment should incorporate an intervention that results in a clear distinction between design teams working with and without hardware.

R3 aims to ensure that H3 is tested by requiring experimental control elements that result in a distinction between design teams working with and without hardware. The rationale behind R3 is to recreate, analyze, and thus, better understand the observed relationship between the question asking behavior of the paper bicycle design team and its use of hardware.

At the beginning of the project, the team did not bring prototyping hardware to its meetings in the design loft, and rarely referenced or examined existing paper bicycles. (Several paper bicycles that were built in the previous years were on display). During these initial meetings, the team operated predominantly at a conceptual level. Approximately halfway through the project, it started building physical prototypes. During this prototyping phase, the team often discovered problems with its designs, and in some cases, appeared to be stuck. It was only then that it began to pay close attention to the bicycles from previous years, examine their design principles, and learn from them.

As outlined in O3 in the previous chapter, initially, when hardware was not present, the questions the team asked were more conceptual and abstract, requiring long answers and leading to detailed discussions. When it started working with prototyping hardware and interacting with the existing artifacts, the questions became considerably more focused and specific.

There might be other causes for the shift in the question asking behavior of the team other than its interaction with hardware. For instance, it is possible that the shift might be a temporal phenomenon related to the life-cycle of a design project. Regardless, H3 focuses on the influence of the access to hardware, and R3 requires the insertion of control elements that replicate the type of interaction the paper bicycle team had with hardware in the experiment.

5.1.3 Design Research Experimentation Related Requirements

R4 through R7 are methodological requirements specific to design research experimentation. In formulating them, I take the position that the main prerequisite of a design experiment—independent of the hypotheses it is attempting to test—is to convincingly simulate a realistic design situation.

R4: The design experiment should promote designing as opposed to problem solving.

In formulating R4, I make a distinction between designing and problem solving, and advocate that the experiment should promote the former. Design researchers often treat designing and problem solving as synonymous parametric processes. It is common to think that what engineers do when they design is to "solve problems."

My position is that although there is truth to this statement, designing and problem solving are fundamentally different. One can choose to view the world—let alone engineering—through a lens which casts most things as problems that need to be solved. This paradigm can be useful if applied selectively. However, if it is overextended, it loses its relevance, and can be rather constraining because there are many situations in life, and in engineering, which require a more open-ended consideration. I believe the term "designing" addresses this very issue by constituting a meta-paradigm, which accounts for problem solving together with other key phenomena such as perception, negotiation, and communication.

More specifically, engineering design theories that are based on the problem solving paradigm assume that design transpires in two distinct domains: the requirements and solutions domains (also referred to as the so-called "requirement" and "solution" spaces). It is also common to assume that the act of mapping the requirement and solution elements contained within the two domains constitute the design activity.

Although I have reservations about subscribing to such an approach, which assumes the existence of requirements and solution domains, I will use it to illustrate my point. Building on existing views regarding the negotiated nature of design requirements [Cuff 1982, Buccarelli 1994, Minnemen 1991, Eodice 2001], I argue that, in a problem solving context, requirements are *given*, and are treated as such by the problem solver, whereas in a design context, they are *negotiated*, and even constructed, by the designer. I also argue that, in a problem solving context, solutions are *final* and take on a static role once formulized, whereas in a design context—borrowing from existentialist thinking—they are constantly *evolving*, never reached, and even never truly exist.

As a simple example, let us consider if the activity an engineering student is engaged in while solving a problem in a statics course—no matter how advanced the course might be—and the activity a practicing design engineer is engaged in while designing a crane are conceptually the same. It is very likely that the engineer and the student will both apply the same theoretical principles and analytical methods in order to analyze and solve "the problem." However, the engineer has to consider and accomplish much more. He/she must consider factors such as why the crane is needed, how and where it will be built, and how and by whom it will be used. He/she must also consider the temporal aspects of such factors: how the needs and usage patterns will change over time.

Therefore, the designer is negotiating and navigating a rich and dynamic situation, whereas the problem solver is solving a bounded and static one. However, the designer will also problem solve when he/she freezes and dissects the dynamic situation, transforms it into static situations, and reduces it into a set of problems. The synthesis of the solutions to the constituent problems informs the designer about the design. However, it does not constitute "the design" as there will always be an indeterminate number of ways of freezing and dissecting any given dynamic situation. A design situation will always yield an arbitrary number of *satisfycing*[10] designs. Therefore, although designing and problem solving are interlinked, they are not conceptually the same thing. R4 formulates the need for this understanding to be incorporated into the design of the experiment.

R5: The setting and scenario of the design experiment should allow for the insertion of control elements associated with the hypotheses without overconstraining the activity (quasi-control as opposed to tight control).

I extend the thinking behind R4 in constructing R5, which requires the experiment to employ quasi-control as opposed to tight control when introducing control elements. Clearly, control elements are needed if the experiment is to qualify as an intervention. However, the nature of the control elements, and hence, the extent of control the experimenter has over the experiment, influences the nature of the activity that will occur in the experiment.

More specifically, in a design context, tightly controlled experiments use interventions and scenarios which aim to test a specific phenomenon. In doing so, they inevitably promote something other than designing—often problem solving—since they force the scenario to point only at the phenomenon, and the activity to revolve around a specific issue, which is usually labeled as "the problem." However, as I argued earlier, designing does not revolve around a singular issue or a problem. Therefore, tightly controlled design experiments fail to simulate realistic design situations, and do not promote design activity.

R6: The design experiment should facilitate the testing of all hypotheses in a single experiment.

R6 requires the design experiment to facilitate the testing of all hypotheses in a single experiment. There are two rationales behind this objective. The first is pragmatic as testing all hypotheses in a single session significantly minimizes the logistical effort required to execute the experiment and the analytical effort to analyze data. The second is related to the distinction between problem solving and designing. If the hypotheses are tested individually in separate sessions, the activity runs the risk of being reduced to fragmented episodes of problem solving, and R4 and R5 cannot be satisfied.

---

[10] The term "satisfycing" is borrowed from Simon [Simon 1981].

On the other hand, testing all hypotheses in a single session can make it difficult to distinguish the phenomena associated with the hypotheses from each other as they might, and most likely would, occur simultaneously. However, that risk can be minimized by the development of clear definitions and metrics for the phenomena in R1, R2, and R3.

R7: The data collection methods used in the design experiment should result in data that can be analyzed qualitatively as well as quantitatively.

R7 ensures that the data generated from the experiments will lend themselves to the analysis techniques that are necessary for testing H1, H2, and H3. Judging from the nature of the phenomena under investigation, it is clear that testing H1 relies more on qualitative techniques, whereas H2 and H3 rely more on quantitative techniques.

The two techniques are fundamentally different in the sense that they require the tracking and measurement of different types of variables. In empirical design research, quantitative techniques require precision in the identification of localized phenomena and repeatability of observation of a given data set in order to account for quantifiable data variables, whereas qualitative techniques require bandwidth of observation in order to capture multiple dimensions of activity and account for potential relationships between qualitative data variables and other related phenomena.

It is necessary to distinguish this point from the distinction I made between ethnographic and audiovisual data collection methods in sections 4.3 and 4.4.1. Although data generated by audiovisual data collection methods are likely to lend themselves to quantitative analysis techniques, they can still be analyzed with qualitative techniques. Similarly, although data generated by ethnographic data collection methods are likely to lend themselves to qualitative analysis techniques, they can still be analyzed with quantitative techniques. In other words, the choice of analysis method is not directly contingent on the data collection method used.

The choice depends on the specifics of the research project and the nature of the data variables. For instance, when conducting field research in order to generate hypotheses, as argued in sections 4.3 and 4.4.1, it is desirable to use both data collection methods and apply qualitative analysis techniques. When testing hypotheses in the laboratory that require the tracking of qualitative as well as quantitative data variables—as is the case with the experiment discussed in this chapter—it is more desirable (and pragmatic) to use the audiovisual data collection method and apply quantitative as well as qualitative analysis techniques.

5.2 Addressing the Requirements

In this section, I will address the requirements discussed in section 5.1, and propose ways of satisfying them in the design experiment.

5.2.1 Defining the Phenomena Outlined in the Hypotheses: The Data Analysis Framework

Developing working definitions for the phenomena outlined in the hypotheses—question timing (hence frequency), question type, design phase, and design team performance—results in an analysis framework for processing the data that will be collected during the experiment, and addresses R2.

5.2.1.1 Question Definition and Type

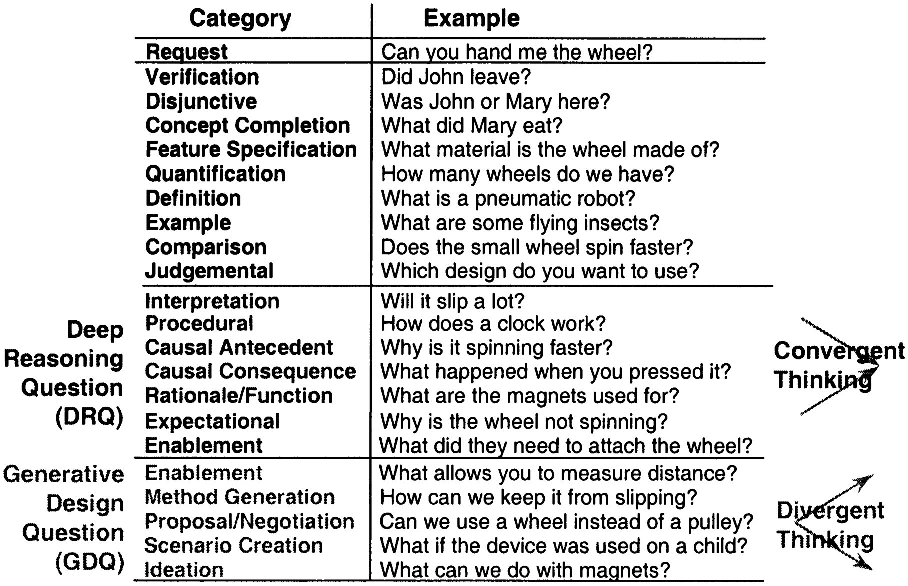
In section 3.2, for the purpose of this study, a question was defined to be a verbal utterance related to the design tasks at hand which demand explicit verbal and/or nonverbal responses. It is important to note that a response constitutes an answer if it has been solicited by the person whose utterance triggered it—responses that were not explicitly solicited do not constitute answers. Otherwise, any verbal exchange would constitute a question-answer pair.

The categories of the taxonomy proposed in section 3.3 can serve as a categorization scheme to determine question type. The final version of the framework, which I based on Lehnert's original question categories, has 22 conceptual question categories—including 4 of Graesser's 5 additions[11], and the 5 Generative Design Question categories I proposed. Therefore, identified questions can be classified according to the 22 categories during the analysis.

The distinction between questions that reflect convergent and divergent thinking constitutes a second classification method (see section 3.3 for a detailed discussion). This method collapses the 22 categories into 3 conceptual classes: Deep Reasoning Questions, Generative Design Questions, and other (Figure 5-1).

---

[11] I did not consider the "Assertion" category Graesser proposed to be a question since the working definition of a question used in this study requires a question to demand an explicit response. An assertion does not necessarily and explicitly seek a response.



	Category	Example	
	Request	Can you hand me the wheel?	
	Verification	Did John leave?	
	Disjunctive	Was John or Mary here?	
	Concept Completion	What did Mary eat?	
	Feature Specification	What material is the wheel made of?	
	Quantification	How many wheels do we have?	
	Definition	What is a pneumatic robot?	
	Example	What are some flying insects?	
	Comparison	Does the small wheel spin faster?	
	Judgemental	Which design do you want to use?	
Deep Reasoning Question (DRQ)	**Interpretation**	Will it slip a lot?	**Convergent Thinking**
	Procedural	How does a clock work?	
	Causal Antecedent	Why is it spinning faster?	
	Causal Consequence	What happened when you pressed it?	
	Rationale/Function	What are the magnets used for?	
	Expectational	Why is the wheel not spinning?	
	Enablement	What did they need to attach the wheel?	
Generative Design Question (GDQ)	**Enablement**	What allows you to measure distance?	**Divergent Thinking**
	Method Generation	How can we keep it from slipping?	
	Proposal/Negotiation	Can we use a wheel instead of a pulley?	
	Scenario Creation	What if the device was used on a child?	
	Ideation	What can we do with magnets?	

Figure 5-1. A conceptual framework of questions based on Lehnert's taxonomy—including 4 of Graesser's 5, and Eris's 5 additional categories. Graesser has termed the Deep Reasoning class. Eris has constructed and termed the Generative Design Questions class, and proposed the convergent-divergent thinking distinction.

Clearly, the second method is simpler, and yet, just as meaningful as the first. Perhaps, it is even more powerful. The finer granularity of the first method can play a descriptive function, whereas the meta-level understanding embodied in the second method can facilitate the testing of the hypotheses.

5.2.1.2 Questioning Rate

In order to determine the questioning rate of design teams during the experiment, all questions should be time stamped. The beginning of the verbal utterance that satisfies the working definition of a question can be taken as the temporal pointer. The rate can be calculated by counting the number of questions asked in one hour, and reported as questions asked per hour. Audiovisual data should be time stamped while recording in order to maintain consistency. This would ensure the existence of a single canonical temporal reference, and free the analysis from device and user dependant variations. The technical aspects of audiovisual recording and replay will be discussed in detail in section 5.2.4.

5.2.1.3 Design Phase and Process

A design phase is a distinct interval of a design process during which functionally similar tasks are performed. The existence of three such phases is commonly agreed on although the vocabulary used to express them can differ: conceptualization, implementation, and assessment. Conceptualization involves need finding, requirements definition, and idea generation. Implementation involves specification generation and product realization (prototyping). Assessment involves product and user testing.

However, design teams do not necessarily perform these phases in that sequential order, nor do they perform them only once. Research in industry has shown that, in real-life product development projects, teams perform design phases in varying durations, sequences, and iterations [Hales 1987, McGown 1999]. These variations might be associated with environmental factors, skills and knowledge base of team members, and other project related elements such as duration, budget, etc.

H1 postulates that the differences in the design processes of teams are reflected in the type and frequency of the questions they ask. It is possible to test that claim by:

1. Monitoring the design processes of teams and observing if specific questioning rates and question types are associated with each phase.
2. Comparing the overall understanding of a team's design process gained by observing a design session, or from viewing the corresponding audiovisual data, with the understanding gained by only considering the frequency, type, and content of the questions that were asked.

5.2.1.4 Design Performance Metrics

Using established performance metrics as a benchmark would enable the testing of the phenomena specified in H2, i.e., the relationship between question asking rate and design team performance. In other words, the metric under consideration, the incidence of questions, needs to be cross-validated with one or more proven metrics.

Before identifying benchmark metrics for cross-validation with the proposed metric, it is useful to classify design performance metrics into two categories according to the nature of the phenomena they evaluate: design performance metrics can be based on phenomena that occur within design activity, or they can be based on the outcome of design activity—the resulting design or prototype. This distinction classifies activity-based metrics as "internal," and outcome-based metrics as "external."

Also, it is necessary to note that when measuring performance, I consider the performance of design teams as opposed to the performance of individual designers for two reasons. Firstly, as discussed in detail in section 2.2, design is a socially mediated activity, and therefore, should be studied as such when possible. Secondly, when designers work in teams, their questioning behavior is much more explicit because questions are a natural part of team communication. The implication is that, when observing a team, it would be very difficult, and even irrelevant, to attempt to measure the performance of individual team members.

The significance and accuracy of the two types of design performance metrics outlined in this section depend on their application context. Since internal metrics focus on design activity, it is most appropriate to use them to measure the quality of the processes of design teams. And, since external metrics focus on the products of design activity, it is most appropriate to use them to measure the quality of the resulting designs—physical prototypes, production drawings, system specifications, etc. However, this appropriation does not imply that internal and external metrics are independent since the outcome of design activity is, by definition, contingent on itself (Figure 5-2). Therefore, internal and external metrics can be assumed to yield corresponding measurements[12].

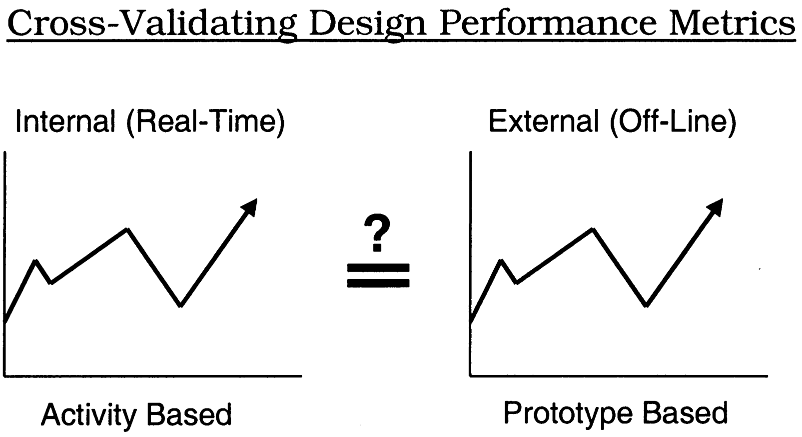


Figure 5-2. The metric under consideration, question asking, needs to be cross-validated with one or more proven metrics. I classify activity based metrics as "internal," and outcome based metrics as "external." The two metrics can be assumed to yield corresponding measurements since the outcome of the design activity is, by definition, contingent on itself.

---

[12] This claim will be revisited and tested during data analysis.

The proposed metric, question asking, is activity based, and therefore, internal. It will be compared with the following two external benchmark metrics for cross-validation:

M1: How well the design satisfies a set of explicit design requirements.
M2: Expert opinion of the quality of the design.

5.2.1.4.1 Benchmark Metric One: Satisfying Given Design Requirements

M1: How well the design satisfies a set of explicit design requirements.

M1 is a measure of how well a design meets its requirements. This metric is appropriate within the context of the experiments since the experimenter will provide the teams with a set of basic requirements. The subjects are still expected to define and negotiate most of the requirements. However, for the purposes of providing structure for the activity and a basis for comparison between teams, a predefined set of requirements will be introduced at the beginning of the exercise.

5.2.1.4.2 Benchmark Metric Two: Experts Judging the Artifact

M2: Expert opinion of the quality of the design.

M2 implies that design performance is, in the case of a multi-user product, a function of how much demand the design ultimately generates from users. This is essentially a measure of how well design requirements might map onto user demands. Experts will be provided with prototypes of the design, and two pieces of key performance information associated with the prototype: price and measurement speed. It is assumed that the average consumer can acquire this information by glancing at the basic specifications listed on the product packaging. Experts will then be asked to reach a judgment based on the provided information and their interaction with the prototype. They will be presented with all of the prototypes, and asked to rank order them as if they are making a purchasing decision.

5.2.2 Intervening to Control Access to Prototyping Hardware

Regulating access to prototyping hardware is one way of promoting a clear distinction between designers working with and without hardware in the experiment. More specifically, half of the teams will be provided with the hardware at the start of the experiment while the other half will be prevented from accessing the hardware until midway through the experiment. The teams that start the exercise with hardware will constitute the control group, and the teams that receive the hardware midway through will constitute the test group. Thus, the intervention will be the delayed introduction of hardware to the test group.

The test teams are expected to conceptualize more in the absence of hardware, and when introduced to the hardware, be more concrete and specific in their thinking. H3 postulates that this will result in an observable change in the types of questions that are being asked. The teams with access to the hardware from the beginning can serve as a control group for comparison. The timing of the introduction of the prototyping hardware will be the control variable.

5.2.3 Promoting "Design Activity" as opposed to "Problem Solving"

R1, R4, R5, and R6 are related; satisfying one implies that the others are satisfied to an extent as well. The relationship between them is expressed in R4, which requires the experiment to promote designing as opposed to problem solving. Therefore, it is appropriate to treat R1, R5, and R6 as subsets of R4.

Deconstructing the experiment into the following two constituents and addressing them separately is an effective way to ensure that the experiment promotes designing as opposed to problem solving: the context in which the exercise takes place, and the scenario.

A team-based (social) environment, which resembles a design setting in industry and requires the subjects to fulfill different organizational functions such as engineering, manufacturing, and marketing, can help establish the appropriate context. This viewpoint is relevant since modern product design is increasingly practiced as an interdisciplinary endeavor, and does not entail individual designers working in isolation. An interdisciplinary approach can sensitize design teams to multiple perspectives and discourage them from taking comfort in a specific domain.

An open-ended design scenario can be utilized in order to guide the teams in the direction of a functional yet novel design. Achieving open-endedness in the design scenario entails defining the endpoint of the design scenario as a direction rather than the comprehension and solution of a specific "problem." The expectation is that an open-ended scenario will encourage the teams to challenge and negotiate the requirements.

5.2.3.1 Employing Quasi-control as opposed to Tight Control

The two methods for addressing the key constituents of the experiment I outlined above—promoting an interdisciplinary approach and defining the endpoint of the design scenario as a direction—also ensure that the experiment will employ quasi-control as opposed to tight control. In other words, they allow for the insertion of control elements associated with the hypotheses without overconstraining the activity.

The analysis framework presented in section 5.2.1 also serves as a means to employ quasi-control. The variables associated with the phenomena that make up the framework occur naturally in design activity, and therefore, can nonintrusively be tracked and measured. The only intrusive control element that can result in a high degree of control over the design activity is the delayed introduction of the prototyping hardware to the test teams. Its effects can be assessed and accounted for by qualitatively comparing the resulting activity of the test teams with the activity of control teams.

5.2.3.2 Testing of all Hypotheses in a Single Experiment

The hypotheses are compatible with each other in the sense that similar design activities need to be observed in order to test them. The hierarchical analytical framework for understanding and measuring design performance presented in section 4.4.4 constitutes evidence for that similarity; the hypotheses build on and complement each other. Therefore, for the purpose of constructing an initial design exercise, it can be assumed that there are no foreseeable obstacles to testing all hypotheses in a single experiment.

The analysis framework presented in section 5.2.1 is specific enough to allow for the accurate identification and tracking of the research variables, which might be occurring simultaneously if all hypotheses are tested in a single experiment.

5.2.3.3 Promoting Realistic Question Asking

For the most part, what I discussed in the preceding sections should ensure that the question asking behaviors of the teams are realistic. In other words, if I can ensure that experiment promotes designing as opposed to problem solving by realizing what I have suggested, it would be plausible to assume that it also promotes realistic question asking behavior.

5.2.3.4 Limitations to Creating Realistic Design Situations in the Laboratory

Attempting to create a realistic design situation in the laboratory has several limitations. This approach should be treated as a "simulation," which implies that the findings can be strengthened by validation in industry.

There are two fundamental limitations: the duration and context of design activity that can be experienced in the laboratory. In the next chapter, I will discuss some additional limitations when I evaluate the nature of the design activity the pilot experiments resulted in.

The duration of a design project in industry can range from weeks to years. The key implication is that designers go through a different learning experience in a longer project. It is likely that the type of learning that takes place over a longer duration influences the nature and frequency of the questions that are asked, and that such influences cannot be accounted for in the laboratory.

The same thinking is valid for the context of the design activity; a laboratory experiment—no matter how complex it might be—provides limited context for a design project, which can only resemble the context of a design project in industry. A conclusive test would need to be carried out in industry for validation.

5.2.4 The Design Observatory: A Research Instrument and Methodology for Capturing Design Activity in the Laboratory

In the laboratory, when testing hypotheses that require the tracking of qualitative as well as quantitative data variables, the most appropriate data collection method is audiovisual recording.

Audiovisual recording provides the necessary precision for identifying localized phenomena and repeatedly observing an existing data set, which quantitative techniques require in order to measure quantifiable variables. It also provides the necessary bandwidth for observation, which qualitative techniques require in order to capture multiple aspects of activity and account for relationships between variables and other related phenomena.

Tang proposed an experimental setting that facilitates the collection of audiovisual data during design activity [Tang 1991]. His configuration evolved over the process of conducting eight design experiments. He advocated that it is beneficial to:

1. Locate the experimenter in a separate room than the room designers are working in.
2. Record multiple views of the design activity.
3. Keep the cameras stationary.

Tang's experimental configuration[13] is illustrated in Figure 5-3.

---

[13] Tang's laboratory was temporary and dismantled after the completion of his research.

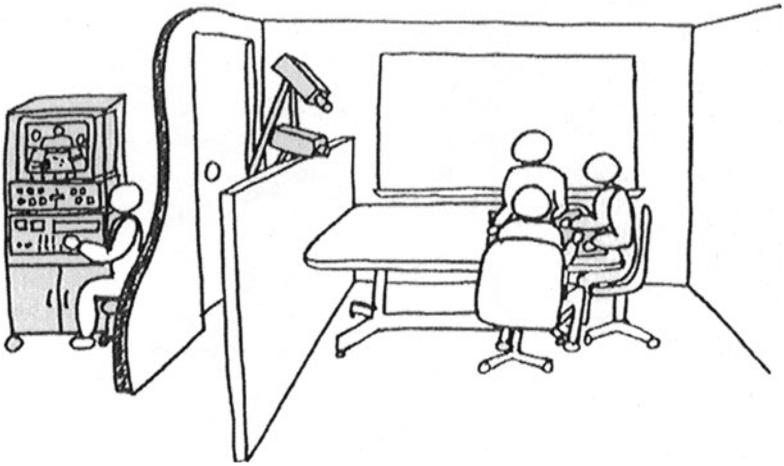


Figure 5-3. Tang's configuration for capturing design activity in the laboratory [Tang 1991]. The experimenter is located in a separate room than the designers. The activity is recorded via multiple stationary cameras.

In order to facilitate high quality audiovisual data collection, and satisfy R7, I decided to build a design research laboratory that would be based on and extend Tang's approach. Together with my design researcher colleagues Carizossa, Milne, and Mabogunje, I undertook the project in November 2000. The resulting space, named, "The Design Observatory," was completed in February 2001.

Similar to Tang's temporary laboratory, the Design Observatory consists of two rooms. One room constitutes the design space where designers—subjects—work. The other room constitutes the data collection and analysis space where researchers monitor experiments and collect and analyze data. In the design space, there are six cameras (mounted at different positions on the walls and the ceiling of the room), five microphones (one is mounted on the ceiling and four are wireless microphones that subjects can use individually), a large whiteboard, a round table, and chairs (Figure 5-4).

Figure 5-4. The design space of the Design Observatory at the Center for Design Research in Stanford University.

In the data collection and analysis space, there is an equipment rack with personal computers that process the audiovisual feeds, a video-quad, an audio-mixer, a television, and a VCR (Figures 5-5). In order to share the specifications of the Design Observatory with the community and aid other researchers who might be interested in building a similar space, we documented the facility in detail in a publication [Carizossa et. all 2002].

Figure 5-5. The data collection and analysis space of the Design Observatory at the Center for Design Research in Stanford University.

During data collection, the experimenter chooses and orients up to four of the cameras prior to experiment, informs subjects of their confidentiality rights, starts the audiovisual recording instruments, introduces the design exercise, moves to the data collection and analysis space, and monitors the experiment and data recording process from there. The resulting audiovisual data are recorded in split screen format, and if four cameras are used, appear in a similar format to the sample frame shown in Figure 5-6.

Figure 5-6. A 4-camera split screen frame from digital video data collected during one of the pilot runs of the design experiment at the Design Observatory.

5.2.4.1 On Collecting and Analyzing Digital Audiovisual Data

Technologically, the most significant contribution of the Design Observatory is its digital media capability; audiovisual data are captured, recorded, and stored in digital format[14]. In that sense, the facility is a technologically enhanced version of Tang's experimental setting.

In a boarder context, utilizing digital technology to capture design activity is not necessarily a new approach. Researchers developing concurrent and collaborative engineering support tools have been, and currently are experimenting with such technologies.

---

[14] As a backup method, audiovisual data are also recorded in analog format with a VCR.

However, utilizing digital technology to analyze data can be seen as a contribution, as it provides new affordances for design researchers. The most significant ones are enhanced audiovisual quality, portability, and the ability to index data. High audiovisual quality shortens analysis time and increases precision. Enhanced portability means that data can be shared faster and with a broader audience, allowing it to be collectively interpreted—inter as well as intra research groups[15]. Enhanced potential for indexing of data can lead to the creation of new cross-referencing methods. Yen has already taken advantage of that potential, and made an advance in cross-referencing of tacit information with sketching activity by developing the software tool RECALL [Yen 2000].

5.3 Meeting the Requirements: The Pilot Experiment

The most productive way of integrating the specifications discussed in section 5.2 into the initial design for the experiment is to review and adopt existing design exercises used by design instructors and researchers that have similar specifications.

The rationale for this approach is embedded in the nature of designing. Since designing is meant to be complex, it is difficult to predict if it will result from a given set of specifications. In order to minimize this risk, the most appropriate starting point is to identify an exercise that is known to have successfully simulated design activity, and then modify it as necessary. In other words, a convenient way to design a design exercise for the purposes of this research was to redesign an existing one with known specifications and similar desired consequences.

With that understanding, I reviewed several existing design exercises. I identified the “Bodiometer Challenge,” originally created by Professor Mark Cutkosky at the Stanford Mechanical Engineering Department, as a suitable candidate. In light of the seven requirements, I modified it to the following form, and used it in the pilot version of the experiment (for the subject instructions provided to the test teams, see Appendix A):

The subjects were asked to design and prototype a measurement device, a "bodiometer," which can be moved along male and female body contours to measure their length, with an operating range from 3 to 100 inches. They worked in teams of three, and had 75 minutes to design and construct a

---

[15] A research project, known as the Delft Protocol Analysis, involving collective interpretation of a data set collected during a design experiment was undertaken by Cross, Christiaans, and Dorst [Cross, Christiaans, and Dorst 1996]. However, data was shared in analog format.

prototype from a standard LEGO parts kit that contained a variety of structural and mechanical components, fittings, and gears. Half of the teams, which formed the control group, were provided with the prototyping materials at the beginning, and the other half, which formed the test group, approximately 35 minutes into the exercise. At the beginning, the test teams received pictures of a representative set of parts that are in the kit instead of the actual hardware (for the pictures that were provided to the test teams, see Appendix B). All teams were provided with a set of instructions and a points scheme, which outlined how their prototype would be scored once it was constructed. The points scheme accounted for performance dimensions such as manufacturability, accuracy, cost, and aesthetics.

Chapter 6

LEARNING FROM THE PILOT EXPERIMENTS: "GOOD" QUESTIONS AND DISCOVERIES

The third step of the empirical dimension of this research has two parts. This chapter addresses the first part, which entails evaluating and redesigning the pilot version of the experiment. The second part will be addressed in Chapter 7. Two pilot runs were conducted, one under the control conditions, and the other under the test conditions. They played a critical role in improving the experimental methodology, deepening my understanding of the nature of questions, and augmenting the hypotheses.

In the following three sections, I assess the implementation of the requirements discussed in the previous chapter in the context of the observations I made during the pilot runs. I also outline how that consideration led to the advancements mentioned above. In the last section, I summarize the augmented hypotheses.

6.1 Improving the Experimental Methodology

In order to improve the experimental methodology, I observed and evaluated the pilot runs with regards to the four design requirements under the design research experimentation criteria, R4 through R7 (for a description of the requirements, see section 5.1).

The pilot runs did not reveal any fundamental difficulties in meeting R4 and R5. As intended, the exercise promoted designing rather than problem solving. The two design teams spent a significant amount of their time and energy in negotiating and redefining the requirements, and explored a wide range of design concepts. For the most part, their approach did not suggest that they viewed the requirements as "givens," and the outcome of their effort as "the solution." They seemed to be aware that both the requirements they were acting on and the designs they were producing were possibilities.

Also, both teams displayed sensitivity to multiple perspectives: they considered user needs, manufacturability and cost issues, and aesthetic values, as well as addressing conceptual and technical issues within their mechanical engineering expertise.

The intervention, delaying the introduction of the hardware to the test team, did not seem to break-up the team's workflow and fragment the activity. The team continued to work without interruption, and did not feel the need to rethink its process when it received the hardware. However, as intended, the intervention influenced the activity by promoting the team to conceptualize more in the absence of hardware. This observation indicates that the nature of the intervention was balanced and not opposed to the natural design process of the team.

However, the pilot runs were instrumental in identifying a number of issues related to R4 and R5. The most significant issue was the timing of the introduction of the hardware to the test group. At the beginning of the exercise, the test group was informed that it would be receiving the hardware 35 minutes after the start of the exercise. During approximately the first 10 minutes, it seemed cognizant of that milestone, but once it got into the exercise and focused solely on designing, it lost track of it. After about 25 minutes, it stopped conceptualizing and indicated that it was ready for the hardware. If it had not lost track of the milestone, it might have paced itself accordingly. I saw no reason to force it to conceptualize for another 10 minutes. Insisting on the intervention in that manner might have interrupted its workflow, so I decided to provide the hardware earlier in the exercise.

In other words, releasing control of the timing of hardware introduction to the team resulted in a smoother transition, improving its workflow. Therefore, in the final runs, I decided that a better way of implementing the intervention would be to give the test teams the choice of asking for the hardware when they felt ready to proceed rather than forcing them to conceptualize for a fixed amount of time.

The pilot runs also revealed that it was necessary to change the structure of the points scheme used for evaluating the prototype according to M1. This modification was necessary to prevent the teams that might be inclined to approach the exercise with a problem solving framework from focusing solely on optimizing their score (the points scheme is outlined in section 7.1.4.1). The intent of the points scheme was to provide the teams with a sense of what might be important to the users of the bodiometer device. However, during the pilot runs, it became clear that when the points scheme was too explicit, it lost its intended function, and instead promoted such teams to be overly concerned with the optimization of the algorithms used for the calculation of the score without considering their meaning.

In the pilot experiment, points could be earned by satisfying each of the following functional and user requirements: accuracy, aesthetics, operation time, number of parts, manufacturing time, and an automated read-out. (For a detailed description of the requirements, see the subject instructions in Appendix A.) The instructions outlined the linear algorithms used in the score calculations. For example, each part used and second elapsed during manufacturing cost the team a fixed number of points. That method resulted in an absolute points scale. Both pilot groups spent significant amounts of time attempting to optimize the relationships between the algorithms in order to maximize their score without considering the intent of the scale.

Therefore, I decided to use a relative points scheme in which points would be assigned based on the rank a prototype achieves among all prototypes in meeting a specific requirement. The teams would not be informed of the performance of other prototypes, and in the absence of that information, be encouraged to consider the meaning of a specified requirement as opposed to calculate the optimal method of satisfying it.

Also, the duration of the exercise proved to be too short for the teams to create a direction for their designs and execute it, as both teams were still negotiating the requirements with 30 minutes remaining in the exercise. Therefore, I decided to increase the duration of the final version of the experiment from 60 to 90 minutes.

Even then, the time limitation had implications. Perhaps, it was the most significant limitation of the experiment since a 90-minute design exercise can never truly substitute for a long-term design project. For example, it is possible that the nature of questions asked by designers change after months of reflection on a design—the taxonomy I use might not even have a category to accommodate such questions. Although I took many steps to ensure that the key characteristics of the questioning behavior of professional designers working on real-life design projects will be replicated in the experiment, I cannot know how successful I have been in achieving that goal unless I attempt to validate my laboratory findings in industry. That is the inverse of what I attempt to accomplish in this research, and would constitute an interesting follow up study.

The pilot runs did not reveal any difficulties in meeting R6, even though testing all hypotheses in the same exercise resulted in the phenomena associated with the hypotheses to occur simultaneously. The definitions I developed for the phenomena, and their expected manifestations in the data, allowed me to identify the research variables and track them independently.

Satisfying R7 by utilizing the digital observation and analysis technology I developed proved to be feasible as well. However, there were two technical issues that needed to be addressed: limitations in mobile digital storage space and playback bandwidth.

I determined that the computer dedicated to capture and playback the audiovisual data needed to support a minimum data transfer rate of 1200 Kb/s in order to attain reasonable image quality at a resolution of 640 by 480 pixels and mono sound at 11.2 kHz scan frequency. The size of a video file captured during a single experiment would be roughly 4 GB. At the time, that was an issue as available portable storage devices such as CD-Roms and floppy discs could not store that much data, and most external hard-drives could not support the 1200 Kb/s transfer rate[16]. However, soon after the pilot experiments, external hard-drives utilizing the FireWire data transfer protocol became available. That technology met the data transfer rate requirement, enabling 15 experiments to be recorded on a single 60 GB external drive.

Storage technology has continued to advance. It is now possible to use DVD-R drives to write digital data onto DVDs that can hold up to 4 GB data each. Thus, audiovisual data from a single experiment can be stored on a single disc. This makes the sharing of digitized experiment data rather effortless as DVDs can be easily replicated and distributed. Also, there are more efficient audiovisual compression protocols available, which should reduce the 4 GB per experiment storage requirement.

6.2 Augmenting the Hypotheses: Discovery Making as another Internal Performance Metric

In order to refine the hypotheses, I reconsidered them in light of the observations I made during the pilot exercises. Although the limited dataset did not permit me to draw conclusions, my observations enabled me to elaborate on their relevance and validity.

When I reconsidered H1, I discovered that paying attention to the nature and timing of questions asked by the two design teams allowed me to gain a comparative understanding of their question asking process. When viewed from a broader scope, that understanding seemed to suggest a topographic representation of the design activity.

I also found qualitative as well quantitative preliminary evidence in the data suggesting that, as postulated in H3, the intervention employed in the experiment affected the questioning behavior of the teams. For instance, the test team asked more questions in the absence of prototyping hardware (a

---

[16] The ability to use portable storage devices is important since one of the main affordances of digital technology is the sharing of data.

21% increase in the second phase of the experiment), whereas the control team asked about the same number of questions in each phase (a 5% increase in the second phase of the experiment). Moreover, the questions asked by the test team in the absence of hardware seemed more conceptual.

Reconsidering H2 raised two issues regarding M1 and M2, the external benchmark performance metrics outlined in section 5.2.1.4. As discussed earlier in this chapter, it was evident that the points scheme used to score the prototypes, the method for obtaining M1, required modification. Even if the points scheme had been sound, comparing the two data points obtained from the pilot runs (M1 results in one performance measurement per team) would not have been meaningful.

It was also evident that obtaining M2, evaluation of the prototypes by experts, was not feasible at that stage for the same reason; experts comparing and ranking only two prototypes was not particularly insightful as a real-time performance measure. Therefore, in the context of the data generated from the pilot runs, it is not meaningful to speculate on the relationship between question asking and the benchmark performance metrics.

Recognizing these issues helped me to identify a characteristic limitation associated with external metrics: measuring performance in terms of the outcome of the design activity, the design, means that the measurement is made on a single object, the prototype, regardless of how many different metrics might be employed. For instance, M1 and M2 are different metrics, but they operate on and judge the same prototype[17].

However, internal metrics are not necessarily subjected to the same limitation since the phenomenon associated with an internal metric most likely occurs numerous times within the activity[18], and it is very possible that each occurrence directly or indirectly causes another performance phenomenon. The identification of a related performance phenomenon might possibly result in another performance metric. Therefore, identifying an additional internal performance phenomenon related to question asking that occurs within the activity would provide multiple measurements, and, hence,

---

[17] The assumption is that there exists a single "design," and hence, prototype. However, even if the outcome of the design activity is considered to be multiple designs, there would be a small number of them. It is unrealistic to think 10 prototypes will be produced in a design project. Although 10 "design concepts" might be created and considered, it is unlikely that more than 3-4 would be implemented in the form of functional prototypes.

[18] If the phenomenon associated with an internal metric does not occur multiple times within the activity, it would be difficult to measure, and attempting to measure it would not be statistically significant. In other words, it would be meaningless to attempt to establish it as a metric.

multiple data points per team even within the limited pilot experiment data set.

In order to identify such a performance phenomenon, I compared my observations of the pilot runs with my observations of the paper bicycle team. I found an observation on the discovery making process of the paper bicycle design team, O2, particularly relevant to what I noticed in the pilot run data. O2 states that the paper bicycle design team seemed to discover more when they asked “good” questions. What I observed during the pilot runs was an extension of that observation: the pilot teams seemed to conceptualize more articulate and a greater number of designs when they discovered more concepts and obstacles. Therefore, I decided to consider “discovery making” as another internal performance metric. This constitutes an additional hypothesis, H4, to supplement the three that were listed earlier.

When identifying a discovery within the activity, I looked for instances where the team experienced a realization that lead to a unique and previously unthought-of concept, or obstacle, related to the design they were working on. I identified four areas in which such conceptual leaps could occur within the scope of the bodiometer design exercise: measurement concept, readout concept, mechanism concept, and obstacle recognition. It is appropriate to note that this method is somewhat similar to judging the effectiveness of a brainstorming session based on the quantity of ideas generated. However, discovery making differs from ideation in the sense that it involves a higher and more visible degree of conceptual continuity and progression. Therefore, it is strongly coupled with learning.

To summarize, my deliberations on the limitation of external metrics and the relevance of identifying another internal performance phenomenon yielded an additional hypothesis, H4:

H4: There is a strong correlation between the incidence of discoveries and design team performance. Hence, discovery making can be taken as a performance metric.

6.3 Refining the Hypotheses: Characterization of a "Good" Question

To facilitate a deeper understanding of the nature of questions, I reconsidered the principles and structure of the taxonomy developed in Chapter 3 by testing it as a coding scheme for the questions asked during the pilot runs. I also expanded on the discussion regarding discovery making, and developed a better sense of what a "good" question might be.

When I attempted to code the questions asked during the pilot runs with the taxonomy, I did not experience any indecision when assigning the questions to the categories—provided I had enough time for each assignment and did not lose focus by coding more than 20 questions in a row without resting. As an alternative coding method, I categorized the questions according to the three encompassing question classes discussed in section 5.2.1.1: Graesser's DRQs categories, the GDQ categories I constructed, and the lower order categories. The alternative method yielded a faster and more decisive coding process.

When I used the taxonomy to code the questions, all of the 22 categories received multiple hits[19]. The distribution was not even as the lower order questions occurred the most. The more significant observation was that I utilized all of the categories and did not encounter any questions that could not be categorized.

As I shifted back and forth between the two coding schemes during the analysis, I began to consider if certain types of questions might be of higher quality than others, and what a "good" might be in a design context. The rationale behind H4 suggests a principle to address this issue. Since the paper bicycle design team discovered more when it asked influential questions, and the pilot teams conceptualized more articulate and a greater number of designs when they discovered more, it was natural to ask: How can the questions that lead to discoveries be identified and characterized?

In order to provide an answer, I assumed "good" questions are associated with discovery making, focused on the instances of discovery making in the data, and identified the preceding questions. A significant part of the questions I identified were DRQs and GDQs.

This observation is in agreement with Graesser's rationale for assigning a higher degree of importance to DRQs than the other types of questions. As discussed in section 2.3.3, Graesser argued that DRQs are associated with achieving the higher level learning goals in Bloom's taxonomy of educational objectives [Bloom 1956], and demonstrated that incidence of

---

[19] During the analysis of data collected from the pilot experiments, I acted as the only coder.

DRQs correlate with learning performance in tutoring situations. However, the tutoring situations Graesser studied do not promote the type of learning that occurs in a design context. Therefore, I wondered if GDQs, which are characteristic of design situations, might also be correlated to performance, but within a design context.

This is not to say that DRQs are not related with design performance. On the contrary, there was no reason to think that their incidence would not contribute to a correlation with performance in a design context as well. Therefore, I postulated that, in order to account for a correlation between question asking and design performance, GDQs needed to be considered in conjunction with DRQs, and that they needed to be treated as a pair.

This consideration can be best studied if it is translated into a hypothesis. The most appropriate way to do so is to incorporate its premise into H2—the existing hypothesis regarding the relationship between question asking and performance—by focusing on the DRQ+GDQ pairs as opposed to all types of questions, and testing for a correlation between the combined incidence of DRQs and GDQs and design team performance.

Therefore, I modified H2 to the following:

H2: Two classes of questions, termed Deep Reasoning and Generative Design questions, are related to design team performance. Their combined incidence correlates strongly with design team performance, and can be taken as a performance metric.

This modified hypothesis, together with the new hypothesis presented in the previous section regarding discovery making, reflect two elements of what a "good" question might be in a design context. Another element is related to the content of a question, which is independent of the consequences and structure of a question. To summarize, three elements of a good question can be taken to be its:

1. Semantic structure
2. Consequences
3. Content

Throughout this work, I argue that two classes of questions, DRQs and GDQs, reflect the semantic structure of good questions, and that the posing of good questions often lead to conceptual leaps, or rather, discoveries.

However, the formalization of the third element, the content of a question, is not addressed in this research, and is somewhat problematic because it is strongly associated with the context the question is posed in. Mabogunje considered this dimension in depth, and argued that contents of design questions are manifested in the "noun phrases" used in design documents and are related to their conceptual and linguistic evolution. He demonstrated this relationship by uncovering a correlation between the incidence of noun phrases in design documents and design team performance [Mabogunje 1997].

Finally, it should be noted that the concept of a "good" question—that there is even such a thing—can be disputed (hence my deliberate effort to keep the qualifier "good" in quotes). It can be argued that there is no such thing as a "bad" question, and that one learns by asking any question. However, in a design situation, the notion of having *intent* and aligning one's thinking with that intent has implications; if one is operating under time, cost, and resource constraints, and is goal-driven, the efficient satisfaction of that goal takes precedence. Therefore, in design thinking, it might be plausible to qualify questions that directly contribute to the realization of a goal as better questions than the ones that do not.

6.4 The Augmented Hypotheses

To summarize, the final states of the hypotheses are the following:

H1: Question timing and type are descriptive characteristics of design cognition and process. When the set of questions a design team asks during a design project is considered as a whole, the timing and nature of those questions point at the fundamentals of the knowledge and rationale the team uses for breaking down and structuring the project into design phases. Question timing and type are informative enough to serve as a roadmap to the design thinking and process of the team.

H2: Two classes of questions, termed Deep Reasoning and Generative Design questions, are related to design team performance. Their combined incidence correlates strongly with design team performance, and can be taken as a performance metric.

H3: Question asking behavior of design teams is influenced by their access to hardware. The types of questions design teams ask change when they transition from working in the absence of hardware to working with hardware.

H4: There is a strong correlation between the incidence of discoveries and design team performance. Hence, discovery making can be taken as a performance metric.

Chapter 7

CONDUCTING THE REDESIGNED EXPERIMENT: PUTTING THE QUESTION ASKING ASPECT OF DESIGN COGNITION UNDER THE MICROSCOPE

The second part of the third empirical step of this research involves conducting the redesigned version of the experiment and analyzing the data. After redesigning the exercise and improving the experimental methodology by reflecting on the pilot experiments, I conducted the final version of the exercise with twelve design teams. I then analyzed the data in order to test the four hypotheses.

In the first section of this chapter, I discuss the data collection and analysis procedures. In the second section, I present my analysis. In the third section, I revisit the hypotheses in light of the results.

7.1 Data Collection and Analysis Procedures

In this section, I discuss the data collection procedures used during the experiments and the analysis procedures used to analyze the data.

7.1.1 Subject Recruitment and Design Team Composition

Subjects were recruited in person and by group email messages. The two prerequisites for being a subject in the experiment were to be a currently enrolled student in a mechanical engineering graduate program at Stanford University and to have no prior knowledge of the "Bodiometer" design exercise. The first twelve subjects volunteered while the remaining 24 were paid $20.00 each for their participation.

Subjects were encouraged to apply in groups of three with people they knew so that they would feel comfortable expressing themselves and ask questions freely. The ones who did so were treated as a design team. Four teams were formed this way. Two of those teams were assigned to the test group, and two to the control group.

There were no guidelines for forming the other eight teams; assignment of subjects to teams, and assignment of teams to experimental groups was performed randomly. However, the subjects making up seven of those eight teams knew each other—they had worked together on a class assignment or a research project, or they were a member of the same academic research group. The subjects making up one of the eight teams had not met before.

It is true that forming teams using this method did not control for heterogeneity across teams, but from the viewpoint of measuring team performance, this was not required.

7.1.2 Experimental Procedure

Immediately before the experiment, design team members were introduced to the functionality of the Design Observatory in order to make them comfortable in the setting. Each audiovisual recording device in the design space was explicitly identified and the procedures for handling captured data were explained. Human subjects consent forms were handed out, and team members were given the necessary time to read and understand the material. Upon receiving written consent from all three members, audiovisual recording was started and subject instructions explaining the design exercise were handed out according to the experimental group the team belonged to—test or control. (For subject instructions, see Appendix A.) The experimenter stayed with the team and answered any preliminary questions until all team members indicated that they understood the instructions.

Before the experimenter left the design space, team members were informed that they could say, "Question," and wait for the experimenter if they had any questions about the exercise. The experimenter then moved next door to the data collection space of the Design Observatory and monitored the activity from there by observing the feeds coming into the digital recording equipment from the cameras and microphone in the design space. If there was a question, the experimenter quickly stepped into the design space, answered it, and returned to the data collection space.

Teams in both experimental groups were notified 30 and 10 minutes before the end of the full 90 minutes. Teams in the test group were given the freedom to decide when to stop conceptualizing and start interacting with the prototyping hardware. However, the prototyping hardware was introduced to the test teams even if they had not asked for it if 35 minutes had elapsed. At the beginning of the experiment, the control teams were provided with the prototyping hardware, and the test teams were provided with a document containing 15 photographs documenting part types. (For the parts catalog provided to the test teams, see Appendix B.)

The hardware consisted of the "Lego Technic Star Wars Episode I Battle Droid" kit (Lego kit number 8001), which had 328 prefabricated structural and mechanical components, fittings, and gears. Each team was provided with a new unopened box containing the kit as well as the original manual with instructions for constructing the Star Wars Battle Droid.

At the end of the 90 minutes, teams were asked to conclude their work. Once the exercise was over, they were asked to identify their prototype and explain how it worked. They were then provided with another Lego kit and asked to identify the parts their prototype was made of. When they were ready, they were asked to construct a device identical to their original prototype. There was no limit on the number of team members who could participate in the construction of the second device, and they were allowed to use the original prototype as a guiding model. The construction process was timed and recorded as the "Manufacturing Time." All audiovisual recording equipment was then turned off.

7.1.3 Transcription

Two of the twelve experiments were fully transcribed. The speaker, time stamp marking the start of the utterance, the utterance itself, and any comments outlining relevant behaviors or circumstances not directly reflected in the utterance, were documented on the transcript. (For a sample segment of the transcript of Team 1, see Appendix C.) Inaudible utterances were clearly marked as such. For reasons I will discuss section 7.1.5, the remaining ten experiments were not transcribed.

7.1.4 Scoring and Judging the Prototypes

The prototypes constructed by the teams were evaluated according to two external benchmark performance metrics discussed in section 5.2.1.4.

7.1.4.1 Scoring the Prototypes according to M1

The first benchmark metric, M1, was a function of how well the prototypes met the stated design requirements: aesthetics, measurement speed, measurement accuracy, manufacturing time, number of parts, and measurement display interface.

A combination of the potential cost and sales of the prototype determined the overall team score. The final score for each team was computed by using the following equations:

Score = Sales – Cost
Sales = Design Concept + Aesthetics + Measurement Time – Error
Cost = Number of Parts + Manufacturing Time

Teams received a score for each variable according to the following rules:

- **Design Concept** was a 30-50 sales point bonus for a design that provided an instrumented readout. Instrumented readout was any method which allowed the user to "read off" a measurement by simply looking at the device without making any calculations or looking at any value tables.
- **Aesthetics** was a subjective category (0-10 points), computed by averaging the scores handed out by 3 judges[20]. Opinions were based solely on the prototype. Visual and intellectual aesthetics were the main considerations.
- **Measurement Time** was the cumulative time it took for the experimenter to make the two measurements. Sales points were handed out in the following way (lower time scores higher): 1st=15, 2nd=13, 3rd= 11, 4th= 10, 5th= 8, 6th= 7, 7th=5, 8th=4, 9th=3, 10th=2, 11th=1, 12th=0.
- **Error** was scored (10 points for each inch of error) was calculated as the absolute value of the difference between the two team measurements and the official measurement.
  Team-measurement = Handweb + Head Circumference
  Error = Absolute Value [(team measurement)-(official measurement)]
- **Number of Parts** was the total number of parts used in the prototype. Cost points were handed out in the following way (higher number scores higher): 1st=15, 2nd=13, 3rd= 11, 4th= 10, 5th= 8, 6th= 7, 7th=5, 8th=4, 9th=3, 10th=2, 11th=1, 12th=0.
- **Manufacturing Time** was the time it took the team to rebuild the prototype from an identical and new parts kit after the main part of the experiment was over. Cost points were handed out in the following way (higher time scores higher): 1st=15, 2nd=13, 3rd= 11, 4th= 10, 5th= 8, 6th= 7, 7th=5, 8th=4, 9th=3, 10th=2, 11th=1, 12th=0.

---

[20] Design Concept and Aesthetics points were assigned subjectively by three Stanford Mechanical Engineering professors.

7.1.4.2 Judging the Prototypes According to M2

The second external benchmark metric, M2, entailed three experts subjectively judging the prototype. All three experts were professors in the Design Division of the Mechanical Engineering Department at Stanford University.

The experts were provided with the prototypes and their cost and measurement speed as defined in the previous section. It is assumed that the average consumer can acquire the same information by glancing at the basic specifications listed on the product packaging. The experts then briefly (5-10 minutes) interacted with the prototypes, and rank ordered them.

7.1.5 Question Identification and Logging

Initially, questions were identified from the transcripts by utilizing the working definition of a question presented in section 3.2. However, identifying questions from the transcripts proved to be problematic as they did not provide the necessary context. The grammar used when posing questions in discourse was often misleading. Many of the utterances that conceptually constituted questions were not grammatically structured as such. Therefore, they could not be identified correctly. For instance, it was difficult to determine if the utterance, "This gear attached to the long rod," constituted a question or not by analyzing the information contained in the transcript.

Even if a question was correctly identified from a transcript, it was fairly difficult to categorize it—again, due to the lack of context. For instance, it was almost impossible to determine if the question, "Can you move the wheel?" should be assigned to the Request or to the Proposal category by simply studying the transcript. Furthermore, in some cases, it was difficult to make such judgments even from the audiovisual data; a 2-3 minute interval in which the question had been posed had to be viewed repeatedly for clarification.

After attempting to analyze the first two experiments from transcripts, it was evident that transcripts could not provide the contextual information audiovisual data did. Also, transcripts were not cost-effective as it took approximately 15 hours to transcribe 1 hour of audiovisual data. Therefore, the other ten experiments were not transcribed, and all experiments were studied primarily by analyzing the audiovisual data directly.

All identified questions were logged on a spreadsheet together with the time stamps marking the start of each question (in seconds from the start of the exercise), and the coded identity of the team member asking the question. Each question was assigned a sequential number (column Q). Each team member was also assigned a sequential number for each question he/she asked (columns A, B, or C). Once a category of the question was determined, the corresponding category number was also recorded (column Cat). A sample spreadsheet segment is displayed in Figure 7-1.

Time	Q	A	B	C	Cat	Question
0						[START EXERCISE]
4	1		1		1	So, we're doing phase one?
89	2		2		22	Why don't we make sure we know how readout's going to be graded?
111	3		3		1	We basically need to measure the perimeter of the contour, right?
160	4	1			18	Does it have to have multiple linkages?
176	5		4		22	We'll write it down as a possible idea, right?
179	6		5		6	What do you call that?
181	7	2			1	That would be a really simple idea--one piece, right?
201	8		6		1	And measure how many revolutions?
208	9	3			18	Or, you could just have a string of legos connected like a linkage?
225	10			1	1	Do you know what I'm saying?
262	11		7		6	What do you mean flipping over?
324	12		8		1	Were you thinking about a one that you'd put together?
348	13		9		3	What do you call that thing?
354	14		10		1	And you keep count?
357	15	4			22	Can I draw something like that just to see if you could X?
384	16		11		1	That was the first idea, right?
400	17		12		21	Any more brainstorming ideas?
415	18		13		1	Is it a requirement that it automatically has to give you a value?
439	19		14		9	I wonder if this would count though, just wrap it around and read it off?
510	20	5			10	Do you think that might be more precise?
522	21	6			6	What's error?
527	22		15		1	It seems like, it also needs to be long enough to go around your head, right?
557	23		16		1	Is that what you're saying?
569	24		17		1	This is 11 inches, right?
609	25		18		6	What's X diameter?
614	26			2	1	Your fingers are about 3 inches long, right?

Figure 7-1. A sample spreadsheet segment where questions asked by design team 12 during the experiment were logged.

7.1.6 Question Categorization

All identified questions were coded according to the categories of the taxonomy of questions presented in section 5.2.1.1. There were two issues associated with the coding process: in certain cases it was difficult to comprehend the context of a question even after viewing the audiovisual data several times, and when the context was determined, the conceptual overlap between some of the question categories added a second degree of ambiguity that needed to be resolved.

In order to comprehend the context in which a question was posed, it was necessary to pay specific attention to and interpret the motivation of the questioner, the general direction of the design activity, the present state of the prototype or sketch or any other representation that was being referenced, and any prior exchanges that might have taken place within the group building up to the question.

Ambiguity resulting from the conceptual overlap between some of the question categories in the taxonomy was resolved by identifying all question category principles applicable to the question under consideration, and prioritizing them in order of intent. In general, it can be assumed that the higher order question categories (in Figure 5-1, categories listed at the bottom are of higher order than the categories above them) are conceptually closer to what the questioner intended, and of higher rank. Therefore, when a question is conceptually in agreement with the defining principles of multiple categories, it should be assigned to the category with the highest rank. For instance, most lower order questions are "Verification" questions, and most DRQs are "Judgmental" questions to some degree. According to the guideline presented here, lower order questions were coded as verification questions only if they could not be coded as belonging to another category. Similarly, DRQ categories had priority over the Judgmental category[21].

Reliability testing was done in order to cross-validate the question identification and categorization processes. Two doctoral candidates, a design researcher and a social scientist, served as coders in the cross-validation process. They were not related to this research and had experience with video interaction analysis and coding. Abiding by the working definition of a question presented in section 3.2, the social scientist was exposed to 50 questions which had been posed by two different teams in two continuous data segments. Fourteen of those questions were either DRQs or GDQs. Cross-validation in question identification yielded 98% reliability. When coding the questions according to the 22 categories, the reliability was 0.90% (4 of the 5 disagreements were related to questions which I had assigned to specific DRQ or GDQ categories). Reliability was 98% when she coded according to the three question classes outlined in section 5.2.1.1.

Since the social scientist did not experience any difficulty in categorizing the questions I had assigned to categories other than the DRQ and GDQ categories, the design researcher was asked to code the questions I had identified as being DRQs or GDQs only. He was exposed to 50 DRQs and

---

[21] Graesser also recognized that the version of the taxonomy he used in categorizing questions could be used as a monothetic or polythetic scheme. He observed overlaps between the Verification category and other categories, and between DRQ categories and other categories. He argued for a similar rank hierarchy to the one presented here based on a slightly different rationale, and opted to use a monothetic scheme [Graesser 1994].

GDQs which had been asked by three different teams in five distinct continuous data segments. This yielded 92% reliability.

7.1.7 Discovery Identification and Logging

After all questions were identified and categorized, the audiovisual data were scanned a second time in order to identify the discoveries the design teams made. As defined in section 6.2, a discovery was considered to be a realization that led to a unique and previously unthought-of concept or obstacle. Each identified discovery was assigned to one of the four discovery categorizes specific to the design exercise used in the experiment: measurement concept, readout concept, mechanism concept, and obstacle recognition.

The categorized discoveries were logged for each design team in a spreadsheet indicating the time the discovery was initially communicated verbally within the team (Figure 7-2, column Time), and the coded identity of the team member communicating the discovery (columns A, B, or C). Since discovery making is a continuous and cumulative phenomenon, it was not appropriate to assign a specific discovery to a specific team member. An aspect of discovery making that could be observed was its initial verbalization.

Each discovery was also labeled with a few descriptive words. The descriptive labels were initially unique to the teams. However, after the discoveries made by all of the teams were logged, similarities emerged between some of them, and the conceptually identical entries were merged under a single discovery label. A spreadsheet outlining the discoveries design team 3 made during the experiment can be seen in Figure 7-2.

7.1.8 Design Phase and Process Observations

As proposed in section 5.2.1.3, design processes of the teams were observed qualitatively while conducting the experiments and analyzing the audiovisual data. Special attention was paid to the sequence and duration of the design phases, and the timing and nature of the questions that were asked.

Although the design phase definitions presented in section 5.2.1.3 and the conceptual question categories of the taxonomy provided structure for the observations, the activity was not strictly reduced to specific analysis units to ensure a holistic approach. Therefore, when investigating the relationship between design process and question asking, design processes of the teams were not formally "coded," but rather evaluated from a broader perspective.

Time	Concept	Readout	Mechanism	Obstacle	A	B	C
114				Can't fit the measurement piece in between fingers	X		
126	Rolling a wheel						X
168		Wheel flicks the read-out					X
179			Gears				X
179			Gear reduction				X
179		Dial					X
324	Series of linkages					X	
420	Exrapolate from a standart body				X		
495	Set Lengths, a					X	
532				Negotiating sharp angled countours and corners	X		
558		Differential			X		
623				Measurement wheel slipping	X		
812			Pulley and rubber bands			X	
1111		Multi-resolution Readout			X		
1402		Slider				X	
1545		Visually count rotations					X
1553		Ticking sound per rotation				X	
1902				Hard to turn at high loads	X		
2840				Doesn't work well on hair	X		
3049				Gears not meshing			X
3634				Low/High Gear Reduction		X	
3786				Rubber bands don't stay on			X
4480				Measurement is not linear			X
4491				Starting position of the wheel effects measurement		X	
4509				Rolling compounds error	X		
4722				Double rubber bands around wheel effects measurement		X	
5050				Dial mark not visible			X
5187				Calibration is invalid if rubber band slips on pulley		X	
5370				A tooth on the dial does not correlate to a rotation			X

Figure 7-2. Spreadsheet outlining the discoveries design team 3 made during the experiment. Time is in seconds.

Taking multiple passes at the data was necessary for gaining that perspective. Each session was observed at least four times. The initial observation was made during data collection, and was continuous. The second and third observations were made during the identification and analysis of questions and discoveries, respectively, and were composed mainly of discrete and shorter sets of observations since the nature of the observations required the observer to pause and review different sections of the data. The final observation was continuous as it was intended to be the final step in obtaining a holistic understanding.

7.2 Data Analysis and Results

I used the analysis procedures presented in the previous section in analyzing the data. Studying the phenomena outlined in the hypotheses lends itself to three fundamental analysis areas: design performance, question asking, and discovery making.

7.2.1 Design Performance

The design performance analysis entailed measuring the performance of each prototype according to the two benchmark metrics, and cross-validating the results.

7.2.1.1 Prototype Performance as Measured by the Benchmark Metrics

I measured the performance of each prototype by applying the procedures outlined in section 7.1.4. The objective performance score (as measured by metric M1) and the subjective ranking associated with each prototype (as measured by metric M2) are displayed in Table 7-1. The prototypes that were ranked higher by the experts were assigned a higher number. The ranking assigned to each team by each expert, as well as the averages of the three rankings of each team, are shown.

Table 7-1. Performance of each prototype measured according to the two external benchmark metrics, M1 and M2. The score and the ranking associated with each prototype are shown. The higher ranked prototypes were assigned a higher number. The ranking assigned to each team by each expert as well as the averages of the rankings for each team are shown. The letter C or T in the team designator indicates if the team belonged to the control or the test group.

Team	M1 (Score)	M2 (Ave. Rank)	Expert 1 Rank	Expert 2 Rank	Expert 3 Rank
1 C	22	4.3	2	9	2
4 C	26	8.7	9	7	10
6 C	11	4.7	5	1	8
8 C	74	12.0	12	12	12
10 C	20	8.3	7	11	7
11 C	49	10.7	11	10	11
2 T	37	6.0	10	2	6
3 T	66	7.0	8	8	5
5 T	31	6.0	3	6	9
7 T	29	4.3	6	4	3
9 T	3	1.7	1	3	1
12 T	22	4.3	4	5	4

7.2.1.2 Cross-validating the Benchmark Metrics

Prior to performing analysis regarding the proposed relationships between question asking, discovery making, and design performance, it is necessary to cross-validate the benchmark performance metrics M1 and M2. If the metrics cannot be cross-validated, findings that might suggest correlation between performance and the phenomena outlined in the second and forth hypotheses cannot be supported with confidence.

Therefore, I performed correlation analysis between the performance values as measured by M1 and M2. The result indicates correlation with high significance (Table 7-2). This finding suggests that the external metrics M1 and M2 are in agreement when they are used to judge the performance of design teams, and constitutes strong evidence for their use as valid benchmarks when testing for the proposed relationships between question asking, discovery making, and design performance.

Table 7-2. Correlation coefficient and significance value obtained by performing correlation analysis between the M1 and M2 performance values for each team presented in Table 7-1.

	R^2	P
Judge Ranking vs. Score	0.55	**0.006**

7.2.2 Question Asking

In this section, I first present the descriptive statistics for the type of questions that were asked during the twelve experiments. I then analyze the proposed relationships between question asking and design process, design performance, and interaction with hardware. Finally, I take a closer look at the interplay between DRQs and GDQs, and demonstrate the relevance of treating them as complementary pairs.

7.2.2.1 Descriptive Statistics of the Types of Questions that were Asked

I analyzed the data on the incidence of questions in conjunction with the results of the question categorization process described in 7.1.6 in producing descriptive statistics for the types of questions that were asked during the 12 experiments. Table 7-3 shows the distribution of the question asking rates among the 22 question categories for each design team.

Table 7-3. Distribution of the question asking rates among the 22 question categories for each design team in questions asked per hour. The letter C or T in the team designator indicates if the team belonged to the control or the test group.

Distribution of Questions among Categories per Team (questions/hr)

	Team Designator											
Question Category	**1 C**	**2 T**	**3 T**	**4 C**	**5 T**	**6 C**	**7 T**	**8 C**	**9 T**	**10 C**	**11 C**	**12 T**
Request/Directive	12.4	9.6	12.1	7.8	6.8	10.7	15.5	19.1	5.2	14.5	13.7	17.6
Verification	48.4	53.1	91.2	68.1	52.5	60.2	57.9	58.8	63.1	50.6	61.6	55.3
Disjunctive	0.6	0.8	3.6	1.2	2.3	2.7	0.6	2.6	1.7	0.7	1.2	2.6
Concept Completion	14.2	8.8	21.8	22.1	9.7	12.7	21.5	9.2	27.5	8.5	13.1	6.5
Feature Specification	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.6	0.0	0.6	0.7
Quantification	3.5	5.5	10.9	8.4	9.1	3.3	6.6	7.9	3.4	3.3	1.9	4.6
Definition	0.6	2.3	1.2	0.6	1.1	0.7	0.6	2.0	0.6	0.0	0.0	2.6
Example	0.0	0.0	0.6	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
Comparison	1.2	0.0	2.4	2.4	2.3	2.0	1.2	0.0	0.6	0.0	0.0	0.0
Judgemental	8.9	4.9	8.5	10.1	8.6	10.7	4.8	10.6	4.6	9.2	1.9	7.2
Interpretation (DRQ)	3.5	2.7	6.6	4.2	5.1	4.0	2.4	5.3	3.4	0.7	4.4	3.3
Procedural (DRQ)	1.2	1.0	0.0	0.0	0.6	2.0	1.8	0.7	1.1	0.7	0.6	1.3
Causal Antecedent (DRQ)	0.0	0.0	0.6	0.6	0.6	0.0	1.2	0.0	0.6	0.0	0.6	0.0
Causal Consequence (DRQ)	0.6	0.0	0.0	0.6	0.6	0.0	0.0	0.0	0.6	0.0	0.0	0.0
Rationale/Function (DRQ)	1.8	4.5	1.8	2.4	5.7	0.0	3.0	3.3	1.1	1.3	3.1	1.3
Expectational (DRQ)	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
Enablement (DRQ)	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
Enablement (GDQ)	0.6	0.8	1.2	0.6	0.0	0.0	0.6	0.0	0.6	2.6	2.5	0.7
Method Generation (GDQ)	6.5	2.3	8.5	2.4	2.9	5.3	3.0	2.0	3.4	5.9	2.5	2.6
Proposal/Negotiation (GDQ)	11.8	13.7	19.3	14.3	7.4	4.7	16.7	22.5	8.6	13.1	14.9	9.1
Scenario Creation (GDQ)	0.0	0.0	0.0	0.6	0.0	0.7	0.6	0.0	0.0	0.7	0.6	0.0
Ideation (GDQ)	3.5	3.9	1.8	1.2	2.3	2.0	3.6	1.3	1.1	3.3	1.2	2.0
Total Questions	119.9	114.9	192.1	148.1	117.5	122.3	141.6	147.2	128.6	115.0	125.1	117.2
Total DRQ	7.7	9.2	9.1	8.4	12.6	6.7	8.4	11.2	7.5	2.6	9.3	5.9
Total GDQ	22.5	20.8	30.8	19.1	12.6	12.7	24.5	25.8	13.8	25.6	21.8	14.3
Total DRQ+GDQ	30.1	30.0	39.9	27.5	25.1	19.4	32.9	37.0	21.2	28.3	31.1	20.2

Table 7-4 reports the distribution of questions among the categories for each design team as the percentage of the total questions asked.

Table 7-4. Distribution of the questions among the 22 question categories for each design team as the percentage of the total questions asked. The letter C or T in the team designator indicates if the team belonged to the control or the test group.

Distribution of Questions among Categories per Team (% of total questions)

	Team Designator											
Question Category	**1 C**	**2 T**	**3 T**	**4 C**	**5 T**	**6 C**	**7 T**	**8 C**	**9 T**	**10 C**	**11 C**	**12 T**
Request/Directive	10.3	8.3	6.3	5.2	5.8	8.7	11.0	13.0	4.0	12.6	10.9	15.0
Verification	40.4	46.2	47.5	46.0	44.7	49.2	40.9	39.9	49.1	44.0	49.3	47.2
Disjunctive	0.5	0.7	1.9	0.8	1.9	2.2	0.4	1.8	1.3	0.6	1.0	2.2
Concept Completion	11.8	7.7	11.3	14.9	8.3	10.4	15.2	6.3	21.4	7.4	10.4	5.6
Feature Specification	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.4	0.0	0.5	0.6
Quantification	3.0	4.8	5.7	5.6	7.8	2.7	4.6	5.4	2.7	2.9	1.5	3.9
Definition	0.5	2.0	0.6	0.4	1.0	0.5	0.4	1.3	0.4	0.0	0.0	2.2
Example	0.0	0.0	0.3	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
Comparison	1.0	0.0	1.3	1.6	1.9	1.6	0.8	0.0	0.4	0.0	0.0	0.0
Judgemental	7.4	4.3	4.4	6.9	7.3	8.7	3.4	7.2	3.6	8.0	1.5	6.1
Interpretation (DRQ)	3.0	2.4	3.5	2.8	4.4	3.3	1.7	3.6	2.7	0.6	3.5	2.8
Procedural (DRQ)	1.0	0.9	0.0	0.0	0.5	1.6	1.3	0.4	0.9	0.6	0.5	1.1
Causal Antecedent (DRQ)	0.0	0.0	0.3	0.4	0.5	0.0	0.8	0.0	0.4	0.0	0.5	0.0
Causal Consequence (DRQ)	0.5	0.0	0.0	0.4	0.5	0.0	0.0	0.0	0.4	0.0	0.0	0.0
Rationale/Function (DRQ)	1.5	3.9	0.9	1.6	4.9	0.0	2.1	2.2	0.9	1.1	2.5	1.1
Expectational (DRQ)	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
Enablement (DRQ)	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
Enablement (GDQ)	0.5	0.7	0.6	0.4	0.0	0.0	0.4	0.0	0.4	2.3	2.0	0.6
Method Generation (GDQ)	5.4	2.0	4.4	1.6	2.4	4.4	2.1	1.3	2.7	5.1	2.0	2.2
Proposal/Negotiation (GDQ)	9.9	11.9	10.1	9.7	6.3	3.8	11.8	15.2	6.7	11.4	11.9	7.8
Scenario Creation (GDQ)	0.0	0.0	0.0	0.4	0.0	0.5	0.4	0.0	0.0	0.6	0.5	0.0
Ideation (GDQ)	3.0	3.4	0.9	0.8	1.9	1.6	2.5	0.9	0.9	2.9	1.0	1.7
Total DRQ	6.4	8.0	4.7	5.6	10.7	5.5	5.9	7.6	5.8	2.3	7.5	5.0
Total GDQ	18.7	18.1	16.0	12.9	10.7	10.4	17.3	17.5	10.7	22.3	17.4	12.2
Total DRQ+GDQ	25.1	26.1	20.8	18.5	21.4	15.8	23.2	25.1	16.5	24.6	24.9	17.2

Finally, Table 7-5 reports a subset of the results, where only the averages for the control and test groups are considered.

These results indicate that approximately half of the questions were Verification questions. This is not surprising as Verification questions are at the lowest level of the taxonomy and instrumental in establishing common ground. The other types of questions that were asked frequently are the Proposal/Negotiation and Concept Completion questions. The high incidence of Concept Completion questions is not surprising either since they are low level questions. However, the high incidence of Proposal/Negotiation questions is significant since they are GDQs. This finding will be addressed in Chapter 8.

Table 7-5. Distribution of the questions among the 22 question categories for the control and test groups in questions asked per hour and as the percentage of the total questions asked. Only the averages for the control and test groups are considered.

Distribution of Questions among Categories for Control and Test Groups (questions/hr and % of total questions)

	Rate (q/hr)		% of Total	
Question Category	Control	Test	Control	Test
Request/Directive	13.0	11.1	10.1	8.2
Verification	57.9	62.2	44.7	46.0
Disjunctive	1.5	1.9	1.2	1.4
Concept Completion	13.3	16.0	10.3	11.8
Feature Specification	0.1	0.2	0.1	0.2
Quantification	4.7	6.7	3.6	4.9
Definition	0.6	1.4	0.5	1.0
Example	0.0	0.1	0.0	0.1
Comparison	0.9	1.1	0.7	0.8
Judgemental	8.6	6.4	6.6	4.7
Interpretation (DRQ)	3.7	3.9	2.8	2.9
Procedural (DRQ)	0.9	1.0	0.7	0.7
Causal Antecedent (DRQ)	0.2	0.5	0.2	0.4
Causal Consequence (DRQ)	0.2	0.2	0.2	0.1
Rationale/Function (DRQ)	2.0	2.9	1.5	2.2
Expectational (DRQ)	0.0	0.0	0.0	0.0
Enablement (DRQ)	0.0	0.0	0.0	0.0
Enablement (GDQ)	1.1	0.6	0.8	0.5
Method Generation (GDQ)	4.1	3.8	3.2	2.8
Proposal/Negotiation (GDQ)	13.6	12.5	10.5	9.2
Scenario Creation (GDQ)	0.4	0.1	0.3	0.1
Ideation (GDQ)	2.1	2.4	1.6	1.8
Total Questions	129.6	135.3	100.0	100.0
Total DRQ	7.7	8.7	5.9	6.5
Total GDQ	21.2	19.5	16.4	14.4
Total DRQ+GDQ	28.9	28.2	22.3	20.8

When the question asking rates of the control and test groups during the exercise are compared, the results seem strikingly similar. More specifically, there is no statistically significant difference between the averages of the DRQ+GDQ and overall question asking rates of the two groups.

Comparison of the DRQ and total question asking rates obtained from the design exercise with the ones Graesser obtained from tutoring sessions yields the results shown in Table 7-6[22]. GDQ asking rates during tutoring are not reported since Graesser does not explicitly account for them. Also, since Graesser does not make a GDQ distinction, he most likely accounts for the Method Generation category that I account for in the GDQ class in his DRQ class under the Procedure category. Graesser also accounts for the GDQ Enablement category under his DRQ Enablement category. Finally, Graesser

---

[22] Graesser's findings were presented in section 3.4.

does not consider the Interpretation category as a DRQ, whereas I do[23]. The DRQ rates reported in Table 7-6 are adjusted to account for DRQs in the way Graesser does to allow for comparison. However, how Graesser accounts for the other three GDQ classes is not clear. It is possible that those types of questions did not even occur in a tutoring context.

Table 7-6. Comparison of the DRQ and total question asking rates observed during the design experiments with the ones Graesser observed during tutoring sessions (in questions asked per hour). The letter C denotes the control group, and the letter T denotes the test group.

	Designing C	Designing T	Tutoring
Total Questions	129.6	135.3	116.3
Total DRQ	9.1[24]	9.2[24]	19.8
Total DRQ+GDQ	28.9	28.2	n/a

The results reported in Table 7-6 show that more DRQs were asked during the tutoring sessions than the design experiments. Since I have not viewed the data from the tutoring sessions, it is difficult for me to account for the difference. Regardless, one explanation can be provided by assuming that the nature of the tutoring sessions promoted the asking of more DRQs; the student and tutor pairs were expected to "converge" on the specified "subject matter" to be learned, and focused on it. However, in the design exercises, no subject matter was specified, and the designers spent a significant portion of their time in generating ideas and expanding, which resulted in a significant number of GDQs in conjunction with DRQs. I will discuss the notion of treating DRQs and GDQs as complementary pairs in detail in section 7.2.2.5.

7.2.2.2 Question Asking and Design Process

In the next two sections, I will analyze the proposed relationships between question asking and design process by using the two analysis procedures I presented in section 5.2.1.3.

7.2.2.2.1 Question Asking and Design Phase

Monitoring the design processes of the teams in order to determine if specific question asking rates and question types are associated with each design phase produced valuable insights.

All design teams went through the three fundamental design phases—conceptualization, implementation, and assessment—multiple times during

---

[23] These differences were discussed in detail in section 3.4.

[24] These DRQ asking rate are different from the ones shown in Table 7-5 since my designation of DRQ categories are slightly different than Graesser's. The adjusted rate is shown so that DRQs are accounted for the way Graesser does.

the experiment. As expected, they did so in varying durations, sequences, and iterations. Some teams were methodical, especially Team 8, and executed them mainly in the above order. Other teams, such as Team 6, began by implementation, moved to conceptualization, back to implementation, and then to assessment. Some teams became predictable, and once they established a phase sequence, they iterated their process by repeating it. Other teams, such as Team 9, were unpredictable, and went in and out of the phases without repeating a pattern. Some teams spent more time in one phase overall than other phases. For instance, Team 5 spent considerably more time than the other teams in the conceptualization phase. Essentially, these observations are a confirmation of the findings of Hales [Hales 1987].

The more significant observation is that such fundamental similarities and differences in the design processes of teams were reflected in the timing and the nature of the questions they asked. In other words, when monitoring the design processes of the teams, I was able to identify relationships between question asking rate, question type, and design phase. Specifically, the strongly pronounced patterns were:

- Teams relied heavily on GDQs during conceptualization.
- Teams relied heavily on DRQs during assessment and implementation.

This observation is illustrated in detail at the question category level with Table 7-7. What I mean by a team "relying" on a specific class of questions is that a class of questions playing a comparatively more influential role in the team's progress toward meeting its design goals than the other classes of questions. These influences needed to be identified mainly through qualitative evaluation. However, in most cases, the incidence of an influential class of questions was higher than the other classes of questions.

The qualitative understanding that lead to the creation of Table 7-7 was based on the fourth and last pass I made at the data. I began the final interaction analysis with an unpopulated version of the matrix presented in Table 7-7 (containing unchecked cells) for each team. When I observed a specific type of question having a strong influence on the team's progress, I identified the design phase the team was in, and placed a checkmark in the corresponding box in the matrix. I took "progress" as making a discovery, or gaining/generating critical knowledge and information that might lead to a discovery (a detailed discussion on discovery making is provided in section 6.1.2). After populating a matrix for each team, I superimposed all of them, and synthesized the general matrix presented in Table 7-7.

Table 7-7. Observed relationships between question types and design phases in design activity. The strongly pronounced patterns were the teams relying heavily on GDQs during conceptualization, and on DRQs during assessment and implementation phases. ■ denotes the types of questions termed as "Deep Reasoning Questions" by Graesser. ● denotes the types of questions termed as "Generative Design Questions" by Eris.

	Design Phase		
Question Category	Conceptualization	Implementation	Assessment
Request	✓	✓	✓
Verification	✓	✓	✓
Disjunctive	✓	✓	✓
Concept Completion	✓	✓	✓
Feature Specification	✓	✓	✓
Quantification	✓	✓	✓
Comparison	✓	✓	✓
Definition	✓		
Judgmental			✓
Interpretation ■			✓
Procedural ■		✓	✓
Causal Antecedent ■	✓	✓	✓
Causal Consequence ■		✓	✓
Rationale/Function ■		✓	✓
Enablement ●	✓	✓	
Method Generation ●	✓	✓	
Proposal/Negotiation ●	✓	✓	
Scenario Creation ●	✓		
Ideation ●	✓		

In the general matrix, the check marks for each question category represent a relative distribution. For example, if Ideation was checked in six of the team matrixes during Conceptualization, and checked in one or two of the team matrixes during Implementation and Assessment, it was only checked in the general matrix during Conceptualization. Also, three types of questions were not asked at all by any of the teams during the experiments: Example, Expectational, and Enablement (DRQ). That is most likely the result of the limited duration of the design exercise. Since I was not able to make any observations on the impact of those types of questions, they are not accounted for in the general matrix.

The associations illustrated in Table 7-7 can be discussed in terms of the principles behind the question categories. Before addressing the distribution of question types to specific phases, I will reflect on the perceived influence of the first seven question categories in all three phases. The first seven categories were closely associated with communication mechanisms, which were geared toward information exchange and social mediation of the activity. Therefore, it is natural for them to appear to have a similar degree of influence in all three phases; they are too fundamental to be dependent on a specific phase. However, the teams asked slightly more verification questions in the implementation and assessment phases.

Another type of question that had a similar degree of influence in all three phases was the Causal Antecedent question which aims to uncover the state or events that has caused the question concept. This might point at a fundamental reasoning mechanism that designers use in establishing causality. Other reasoning mechanisms that directly address causality are embodied in the Causal Consequence and Rationale/Function questions. However, in order for those questions to have an influential role, concrete events or concepts should have already been constructed. For instance, the Causal Consequence question, "What happened when you pressed it," assumes the existence of an artifact that was operated on. These types of opportunities for asking Causal Consequence and Rationale/Function questions were less likely to occur during conceptualization, where designers were not necessarily concerned with firmly grounding themselves in existing events, concepts, or artifacts.

When the distribution of question types to specific phases was considered, conceptualization and assessment phases had distinct profiles. Conceptualization involves tasks aimed at need finding, requirements definition, and idea generation. Therefore, Definition, Scenario Creation and Ideation questions were influential. The other GDQs—Enablement, Method Generation and Proposal/Negotiation questions—were equally influential during conceptualization phases, however, they did not contribute to the unique profile as they proved to be pivotal during implementation as well.

During assessment, Interpretation and Judgmental questions were instrumental in testing prototypes and determining if they met the requirements. In evaluating prototypes, designers often expressed a need to extrapolate the behavior of the prototypes to realistic situations that involve users. Interpretation questions played a critical role in extending their understanding of prototypes, and Judgmental questions constituted a natural mechanism for initiating and concluding decision making processes.

Implementation phases were rather comprehensive and relied on a wide range of questions. That was mainly due to the transitional nature of implementation tasks, during which designers generated specifications from the needs, the requirements, and the concepts which had been generated and defined during conceptualization. Thus, during implementation, the focus was also on "generation," but it was narrower and goal driven. Therefore, Procedural, Method Generation, Enablement, and Causal Consequence questions were especially influential.

7.2.2.2.2 Comparison of Meta-Level Understandings

I was able to gain an understanding of the design processes of the teams—how they structured their design tasks and reflected that structure in their workflow—while conducting the experiments and viewing the resulting audiovisual data. As an alternative method, I considered only the frequency, type, and content of the questions they asked. Comparison of the understandings gained through these two methods revealed similarities that complement and strengthen the results presented in the previous section.

Most teams explicitly considered breaking down their activity into tasks and proposed a structure for their work. As mentioned in the previous section, some were methodical while others only used the minimum level of structure they thought was necessary.

Teams such as Team 6 did not pay much attention to planning their tasks and improvised as they went along. It can be argued that this team, and others like it, did not have structure, and that their activity should not constitute valid data for design process observations; if the team did not seem to care for structure, what process was there to study? However, what I observed in their work is that the absence of explicit planned structure resulted in emergent structure of a spontaneous nature, and the resulting activity was worthy of consideration for that reason.

When gaining a meta-level understanding of the design process of each team, I paid special attention to a number of descriptive elements of the activity that seemed to be strongly influenced by the design processes of the teams. These were:

1. The local goal the team was working toward at any given time.
2. The general topic(s) of discourse. This was usually dependent on the local goal.
3. Change in the direction of discourse. This was usually triggered by the negotiation of the local goal.
4. Social elements such as leadership, and cognitive and political interplay.
5. The level of cognitive progress. This was reflected in the degree of completion of the team's overall design goal.
6. The rate of change in cognitive progress. This was related to the rate at which the team was making conceptual leaps, or, discoveries, and getting closer to accomplishing its overall design goal.

When there was a change in the process of a design team, or rather, when a team entered a different phase in its design process, that change was usually reflected in these elements. More specifically, elements 1, 2, 3, and 4 were reflected in the questions rather strongly and continuously, and elements 5 and 6 were reflected partially and sporadically. By repeatedly observing such influences during an entire design session, and by doing so for each of the twelve teams, I was able to form an opinion on the design process of each team.

In order to gain a meta-level understanding of the design process of each team through the questions they asked, I reviewed the spreadsheets where the questions were logged. (A sample section of the spreadsheet for Team 12 is illustrated in Figure 7-1.) I read through each spreadsheet a minimum of three times, considering the frequency, type, and content of the questions, and attempted to identify and track the descriptive elements listed above. By synthesizing the elements that could be identified and tracked from the spreadsheets, I constructed a second understanding on the design process of each team. I then compared that understanding with the initial, and more accurate, understanding I gained through direct observation.

After performing this analysis for each team, I concluded that the fundamentals of how a design team structured its design tasks could be reconstructed by analyzing the frequency, type, and content of the questions they asked.

Although this is a significant finding, there are two limitations associated with it. Firstly, the independence of the two understandings I gained of the design process of each team can be questioned since, in order to compare them, I needed to gain one understanding before the other. The insight I thought I gained through analyzing the spreadsheets might have already been with me, acquired while conducting the experiments and viewing the resulting audiovisual data. Since evaluation was qualitative, there is no objective way of refuting that claim. However, I made sure that I performed the two methods independently by allowing for a minimum of two weeks between the time I completed the direct observations and began analyzing the spreadsheets.

Secondly, the understanding of the design processes of a team I gained by analyzing the spreadsheets was rudimentary, and does not constitute an undiminished substitute for the understanding I gained by observing the activity directly; at best, it constitutes a reduced set. However, that is not to say it is not descriptive enough. On the contrary, it would be most appropriate to characterize it as a topographic representation of the design activity, and hence, as a roadmap to the design thinking and process of a team.

7.2.2.3 Question Asking and Performance

Identifying and categorizing the DRQs and the GDQs that were asked during the experiments enabled me to test the proposed relationships between question asking and performance. Prior to focusing on the GDQ-DRQ pairs as suggested in H2 and H3, it is relevant to test for correlation between the overall question asking rates—without making any distinctions between the nature of questions—to ensure there is none. If there is a correlation, focusing on the DRQ-GDQ pairs might not be not as relevant as the hypotheses state.

The combined GDQ+DRQ and overall question asking rates, and the prototype scores for each design team are shown in Table 7-8. The averages for the test and control groups are also shown.

Table 7-8. Combined GDQ+DRQ and overall question asking rates, and prototype scores of each design team. Averages of the test and control groups are shown in the last two columns. Results are reported in questions asked per hour. The letter C or T in the team designator indicates if the team belonged to the control or the test group.

Question Asking Rates and Prototypes Scores per Team and Averages for the Control and Test Groups (questions/hr)

	1 C	2 T	3 T	4 C	5 T	6 C	7 T	8 C	9 T	10 C	11 C	12 T	C	T
Total Questions	119.9	114.9	192.1	148.1	117.5	122.3	141.6	147.2	128.6	115.0	125.1	117.2	129.6	135.3
Total DRQ+GDQ	30.1	30.0	39.9	27.5	25.1	19.4	32.9	37.0	21.2	28.3	31.1	20.2	28.9	28.2
Score	22.2	36.6	65.7	25.7	30.9	11.0	28.7	74.3	2.8	19.7	48.5	21.5	33.6	31.0

There were no statistically significant differences between the averages of the combined DRQ+GDQ and overall question asking rates of the two groups (see section 7.2.2.1). Analysis of the prototype score data shown in Table 7-8 yielded a similar result for the differences between the averages of the scores of the two groups.

When the overall question asking rates of the twelve design teams were plotted against their prototype scores, no correlation was visible (Figure 7-3). Statistical analysis confirmed this observation by yielding weak correlation coefficients with low significance (Table 7-9, row 2).

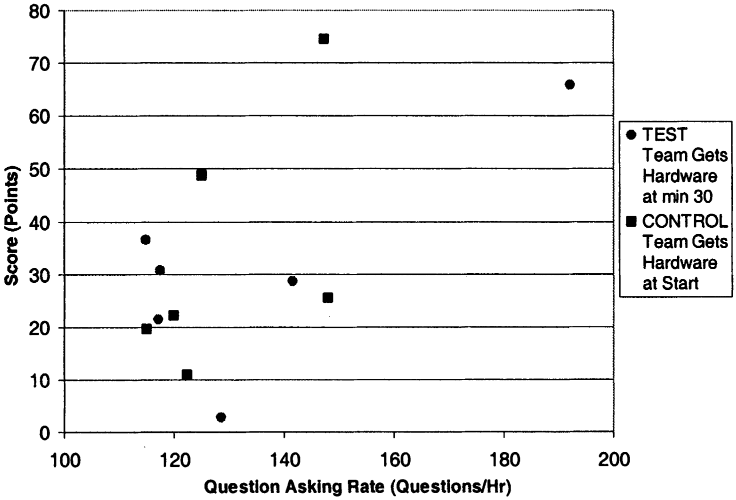


Figure 7-3. Overall question asking rates of the twelve design teams plotted against their prototype scores. Data points marked by squares belong to the teams in the control group, and points marked by circles belong to the teams in the test group.

However, when the combined DRQ+GDQ asking rates of the twelve design teams were plotted against their prototype scores, a linear relationship suggesting positive correlation was visible (Figure 7-4).

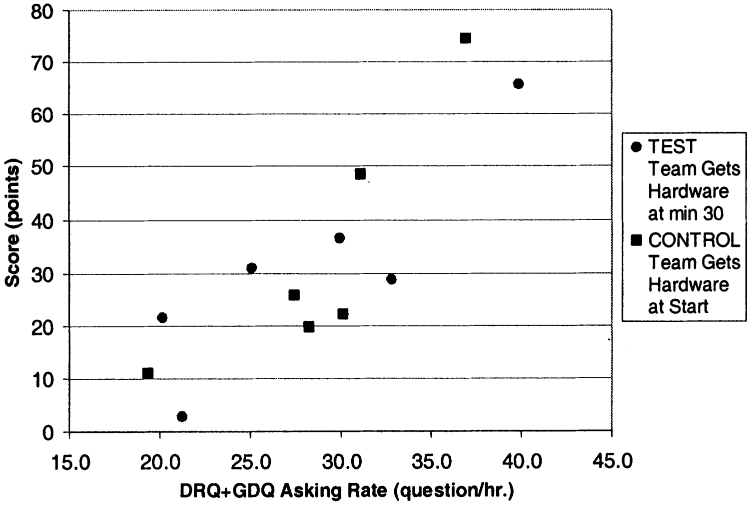


Figure 7-4. Combined DRQ+GDQ asking rates of the twelve design teams plotted against their prototype scores. Data points marked by squares belong to the teams in the control group. Data points marked by circles belong to the teams in the test group.

Statistical analysis of the data plotted in Figure 7-4 yielded strong correlation coefficients with high significance values (Table 7-9, row 1) for both the control and the test groups.

Table 7-9. Correlation coefficients (adjusted R^2) and significance values for correlation between team score and GDQ+DRQ, DRQ, GDQ and overall questions asking rates. Bold numbers indicate strong correlation or high significance. Lighter numbers indicate weaker/no correlation or lower/no significance.

	Control R^2	Test R^2	Control P	Test P
GDQ+DRQ vs. Score	**0.68**	**0.70**	**0.027**	**0.023**
All Questions vs. Score	0.13	0.39	0.260	0.110
DRQ vs. Score	0.45	0.10	**0.087**	0.514
GDQ vs. Score	0.15	**0.56**	0.239	**0.054**

In order to ensure that the incidence of neither DRQs nor GDQs could establish the positive correlation alone, I analyzed the relationships between DRQ and GDQ asking rates and prototype scores for correlation independently. DRQ asking rates of the control teams correlated positively with prototype scores (Table 7-9, row 3). GDQ asking rates of the test teams correlated with prototype scores (Table 7-9, row 4). However, DRQ asking rates of the test teams, and the GDQ asking rates of the control teams did not correlate with the prototype scores. Also, the strength and significance of the correlation between DRQ asking rates of the control team and the GDQ asking rates of the test teams and prototypes scores was much less than the correlation between the combined DRQ+GDQ asking rates of both groups and prototype scores.

These findings demonstrate that DRQs and GDQs need to be treated as *complementary pairs* when it comes to establishing their value as a design performance metric.

7.2.2.4 Question Asking and Interaction with Hardware

Observing the changes in the combined DRQ+GDQ asking rates of the teams in the test group as they transitioned from the initial part of the experiment, Part A, where they were encouraged to conceptualize in the absence of prototyping hardware to the second part of the experiment, Part B, where they were given access to hardware, and comparing those changes to the changes in the combined DRQ+GDQ asking rates of the teams in the control group during the corresponding time intervals yielded the necessary results for evaluating H3. In H3, I postulated that the DRQ+GDQ asking rates of design teams change when they transition from working in the absence of hardware to working with hardware.

The results are striking as the average of the combined DRQ+GDQ asking rate of the teams in the test group decreased by 21% from Part A to Part B, whereas it increased by 3% for the teams in the control group (Figure 7-5).

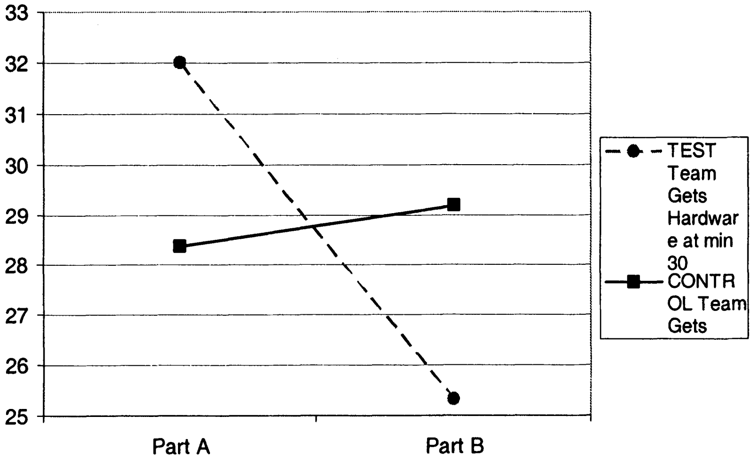


Figure 7-5. Averages of the combined DRQ+GDQ asking rates of the teams in the test and control groups in Parts A and B of the experiment.

The difference between the averages of the combined GDQ+DRQ asking rates for the test group was statistically significant, whereas the difference between the averages for the control group was not (Table 7-10, row 1). Therefore, it can be concluded that the average of the GDQ+DRQ asking rate for the test group decreased significantly, whereas it did not exhibit any meaningful change for the control group between parts A and B of the experiment.

Table 7-10. Significance values for the difference of the average of the combined GDQ+DRQ, GDQ, and DRQ asking rates of the control and test teams between Part A and Part B of the experiment. Bold numbers indicate high significance. Lighter numbers indicate lower/no significance.

	Control P	Test P
Part A vs. Part B — GDQ+DRQ	0.391	**0.063**
Part A vs. Part B — GDQ	0.493	**0.104**
Part A vs. Part B — DRQ	0.286	0.409

Further analysis revealed that the decrease in the average of the combined DRQ+GDQ asking rates of the test teams was directly associated with the decrease in the average of their GDQ asking rates (Figure 7-6) since the averages of their DRQ asking rates did not change significantly (Table 7-10, rows 2 and 3).

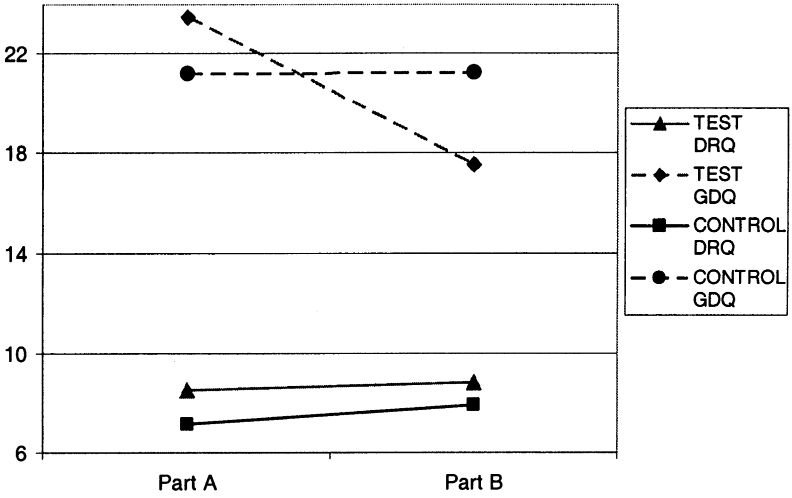


Figure 7-6. Averages of the DRQ and GDQ asking rates of the teams in the test and control groups in Parts A and B of the experiment.

Therefore, the combined GDQ+DRQ asking rates of the design teams in the test group initially working in the absence of prototyping hardware decreased when they transitioned to working with hardware, and that the combined GDQ+DRQ asking rate of the design teams in the control group did not exhibit any significant change between the corresponding time intervals. These findings demonstrate that question asking behavior of design teams is influenced by their access to hardware.

7.2.2.5 DRQs and GDQs as Complementary Pairs

The findings reported in section 7.2.2.3 demonstrate that DRQs and GDQs need to be treated as complementary pairs when it comes to establishing their value as a design performance metric. Based on the data collected within the scope of this research, there are at least three additional analysis methods that can be performed in order to gain a deeper understanding of that relationship.

The first approach would be to hypothesize that there is an optimal DRQ to GDQ asking ratio, and to investigate the relationship between the DRQ/GDQ asking ratios and performance for each team. The second approach would be to hypothesize that there are cyclic relationships between

DRQs and GDQs, to identify the transitions between DRQs and GDQs, and test for correlation between their DRQ-GDQ transition rates and performance. The third approach—the most complex one—would be to hypothesize that there is causality between DRQs and GDQs, and to analyze the data for patterns which might reveal causal links between their occurrences.

At this stage of the research, I only performed the first two approaches. In applying the first method, I calculated the DRQ/GDQ asking ratios for each team, which are reported in Table 7-11, row 1. When the DRQ/GDQ asking ratios are plotted against the prototype scores for each team, an optimal ratio was not visible (Figure 7-7). However, it was clear that 10 of the 12 design teams asked approximately 4 DRQs for every 10 GDQs. Even though this observation does not have any significance in suggesting a relationship between DRQ/GDQ asking ratios and performance, it suggests that 0.4 might be a fundamental DRQ/GDQ ratio in the context of designing.

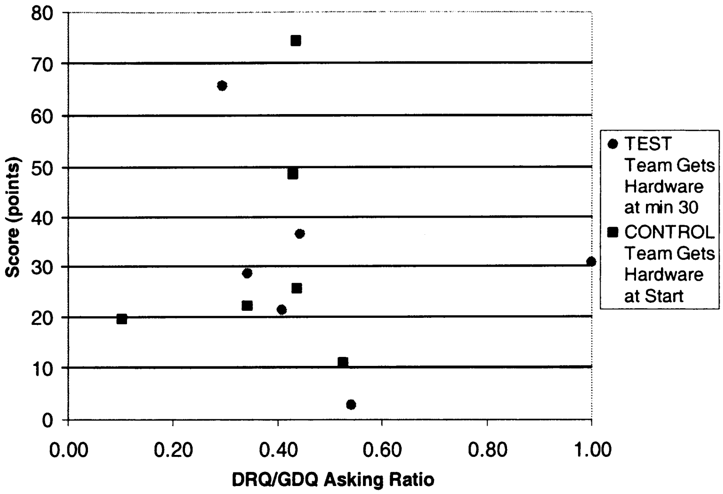


Figure 7-7. DRQ/GDQ asking ratios of the design teams plotted against their prototype scores. Data points marked by squares belong to the teams in the control group. Data points marked by circles belong to the teams in the test group.

In performing the second method, I isolated and considered the data on DRQs and GDQs. I chronologically sorted the DRQs and GDQs each team asked, and accounted for the frequency of the transitions between them. The combined DRQ+GDQ asking rates, the prototype scores, and the DRQ-GDQ transition rates for each design team are shown in Table 7-11. The averages for the test and control groups are also shown.

Table 7-11. Combined GDQ+DRQ asking and DRQ-GDQ transition rates, prototype scores, and DRQ/GDQ ratios of each design team. Averages of the test and control groups are in the last two columns. Results are reported in questions asked and transitions made per hour. The letter C or T in the team designator indicates if the team belonged to the control or the test group.

Combined DRQ+GDQ Asking Rates, Scores and DRQ-GDQ Transitions per Team and Averages for the Control and Test Groups (per hr)

	1 C	2 T	3 T	4 C	5 T	6 C	7 T	8 C	9 T	10 C	11 C	12 T	C	T
DRQ/GDQ Ratio	0.34	0.44	0.29	0.44	1.00	0.53	0.34	0.44	0.54	0.10	0.43	0.41	0.38	0.50
Total DRQ+GDQ	30.1	30.0	39.9	27.5	25.1	19.4	32.9	37.0	21.2	28.3	31.1	20.2	28.9	28.2
Score	22.2	36.6	65.7	25.7	30.9	11.0	28.7	74.3	2.8	19.7	48.5	21.5	33.6	31.0
DRQ-GDQ Tans.	10.0	14.7	14.5	10.7	9.1	6.7	11.3	19.1	9.8	5.3	12.4	9.1	10.7	11.4

Statistical analysis yielded strong correlation of high significance between the DRQ-GDQ transition rates and prototype scores for the control group, but not for the test group (Table 7-12, row 1). The difference between the results of the test and control groups might be related to the behavior illustrated in Figure 7-6—natural transition patterns might have been affected by the intervention.

Table 7-12. Correlation coefficients (adjusted R^2) and significance values for correlation between team DRQ-GDQ transition rate and prototype score, and DRQ, DRQ-GDQ transition and combined DRQ+GDQ asking rates. Bold numbers indicate strong correlation or high significance. Lighter numbers indicates weaker/no correlation or lower/no significance.

	Control R^2	Test R^2	Control P	Test P
DRQ to GDQ Transitions vs. Score	**0.85**	0.41	**0.005**	0.101
DRQ+GDQ Asking vs. Transitions	0.55	0.56	0.055	0.053

When interpreting the strong correlation between the DRQ-GDQ transition rates and prototype scores for the control group, it is necessary to keep in mind that the teams that ask more DRQs+GDQs score higher (Table 7-9, row 1). Therefore, it is also necessary to consider that the teams that ask more DRQs+GDQs will be more likely to execute more DRQ-GDQ transitions. Statistical analysis supports this explanation; there is significant correlation between DRQ-GDQ transition and asking rates (Table 7-12, row 2). More analysis is required to determine the extent the relationship between DRQ+GDQ asking rates and the score might be contributing to the correlation between DRQ-GDQ transitions and score.

Although the results of the two analysis methods discussed in this section do not lead to significant conclusions, they strongly suggest that studying the interplay between the DRQ-GDQ pairs further might be revealing. The third analysis method mentioned would most likely be instrumental in gaining that understanding.

7.2.3 Discovery Making

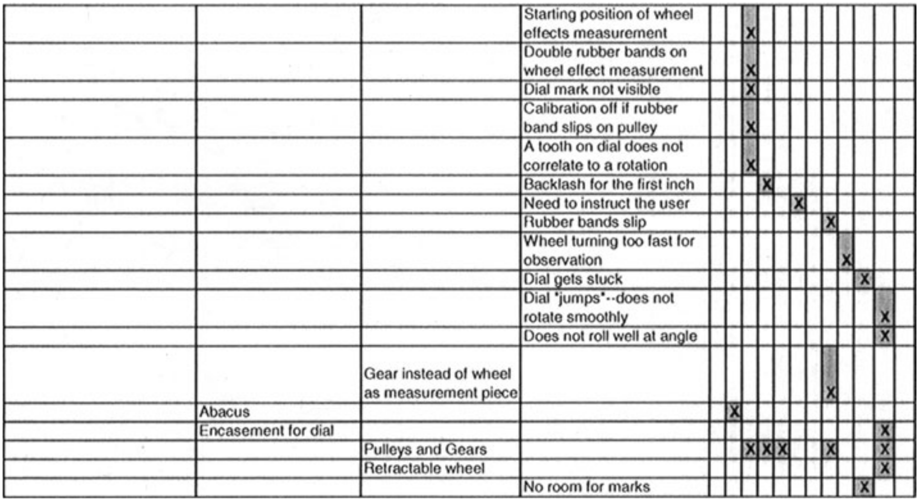
In this section, I present and categorize the discoveries that were made during the experiments, and analyze the relationships between discovery making, question asking, and performance.

7.2.3.1 Categorization and Logging the Discoveries

I identified the discoveries according to the definitions and procedures outlined in sections 7.1.7. After logging the discoveries made by each team in separate spreadsheets as illustrated in Figure 7-2, I merged them into a single spreadsheet where all of the discoveries were accounted for under the four categories (Figure 7-8).

Overall, 38 discoveries were made regarding measurement, readout and mechanism concepts, and 31 discoveries were made regarding obstacles. Qualitative examination of the discoveries reveals that the teams were able to generate ideas that are conceptually distinct and unique despite the limitations of the laboratory setting. Considering that the experiment lasted only 90 minutes, these findings demonstrate that a wide range of discoveries were made—quantitatively and conceptually—and suggest that the experiment was successful in generating design activity as opposed to problem solving.

Measurement Concept	Readout Concept	Mechanism	Obstacle	1	2	3	4	5	6	7	8	9	10	11	12
Rolling a wheel--translating rotation into distance				X	X	X	X	X	X	X	X	X	X	X	X
Exrapolate from a standart body part					X	X	X	X	X	X		X			X
Series of linkages				X		X			X		X				X
"Set Lengths", a Rod				X	X	X		X			X				
Paper-Pencil outline				X	X			X	X						
Interchangable/multiple wheels								X			X	X	X		
Rubber Bands String					X						X			X	
Stationary Device User moves hand												X	X	X	
Calipers										X					X
Ratchet													X	X	
Hand Displacement				X											
Tank Tracks								X							
Separate devices for the measurements									X						
Tight-hat mechanism										X					
Milling machine										X					
Thumb dial										X					
Tweezer											X				
Rolling a noncicular object											X				
Device conforming to hand														X	
Pivoting links traversing the contour															X
	Dial			X	X	X	X	X	X	X	X	X	X	X	X
	Visually count rotations			X	X	X	X	X		X		X		X	
	Multi-resolution			X	X	X	X			X		X	X	X	
	Ticking sound per rotation			X		X	X	X				X	X		X
	Dial rotates twice						X			X	X	X		X	
	Flipping Magnet			X				X							
	Physical Memory (automatic mark /turn)				X	X									
	Slider				X	X									
	Physical Memory (manual mark per turn)				X						X				
	Differential					X									
	Winding rubber band unwinds string							X							
		Gears		X	X	X	X	X	X	X	X	X	X	X	X
		Gear/Pulley Reduction		X	X	X	X	X		X	X	X	X	X	
		Rubber band around measurement wheel		X	X		X	X		X	X	X	X	X	
		Pulley+rubber band		X	X	X	X			X	X				
		Wheel rotates arm which ticks read-out				X	X					X	X	X	
		Eccentric cam					X								
		Rack moving dial											X		
			Measurement wheel slips	X	X	X	X	X	X	X	X	X	X	X	
			Not enough gear ratio		X	X	X	X		X	X	X	X	X	
			Low Resolution	X		X	X	X			X			X	
			Device-User interference	X			X	X		X	X	X			
			Measurement piece not fitting between fingers		X	X			X	X	X	X			
			Doesn't work well on hair		X	X	X	X	X				X	X	
			Negotiating sharp angled countours and corners			X		X	X		X				X
			Too much friction in drive			X	X				X	X			
			Wheel shaft doesn't spin well				X	X		X				X	
			Meshing gears too tight			X		X		X					
			Rolling compounds error			X	X								X
			Too much pulley tension	X									X		
			Gears not meshing	X		X									
			Rubber bands come off		X	X									
			Limited data set for extrapolation						X						X
			Difficult self application						X	X					
			Drive mechanism stuck							X	X				
			Magnetic force too strong	X											
			Nonlinear measurement			X									



			Starting position of wheel effects measurement			X									
			Double rubber bands on wheel effect measurement			X									
			Dial mark not visible			X									
			Calibration off if rubber band slips on pulley			X									
			A tooth on dial does not correlate to a rotation			X									
			Backlash for the first inch				X								
			Need to instruct the user						X						
			Rubber bands slip								X				
			Wheel turning too fast for observation									X			
			Dial gets stuck										X		
			Dial "jumps"--does not rotate smoothly											X	
			Does not roll well at angle											X	
		Gear instead of wheel as measurement piece									X				
	Abacus				X										
	Encasement for dial													X	
		Pulleys and Gears				X	X	X			X			X	
		Retractable wheel												X	
			No room for marks										X		

Figure 7-8. Spreadsheet summarizing all of the discoveries made by the 12 design teams. If a team has made a particular discovery, an "X" appears in the cell under the corresponding team column and across the corresponding discovery row. Otherwise, the cell is left blank. In each category, the discoveries that were made by a larger number of teams are listed higher in the table. Teams 1, 4, 6, 8, 10, and 11are in the control group. The others are in the test group.

7.2.3.2 Discovery Rate and Performance

Identification of the discoveries the design teams made during the experiment provided the necessary insights for testing H4. The discovery rate, the combined DRQ+GDQ asking rate, and the prototype score of each design team are shown in Table 7-13. The averages of the test and control groups are also shown.

Table 7-13. Discovery rate, combined DRQ+GDQ asking rate, and prototype score of each design team. Averages of the test and control groups are shown in the last two columns. Results are reported in discoveries made and questions asked per hour. The letter C or T in the team designator indicates if the team belonged to the control or the test group.

Discovery and Combined DRQ+GDQ Asking Rates, and Scores per Team and Averages for the Control and Test Groups (per hr)

	1 C	2 T	3 T	4 C	5 T	6 C	7 T	8 C	9 T	10 C	11 C	12 T	C	T
Discoveries	11.8	16.4	17.5	14.3	12.6	10.0	13.1	15.8	10.9	11.2	13.1	8.5	12.7	13.2
Score	22.2	36.6	65.7	25.7	30.9	11.0	28.7	74.3	2.8	19.7	48.5	21.5	33.6	31.0
Total DRQ+GDQ	30.1	30.0	39.9	27.5	25.1	19.4	32.9	37.0	21.2	28.3	31.1	20.2	28.9	28.2

When the discovery making rates of the design teams were plotted against their prototype scores, a linear relationship suggesting positive correlation was visible (Figure 7-9).

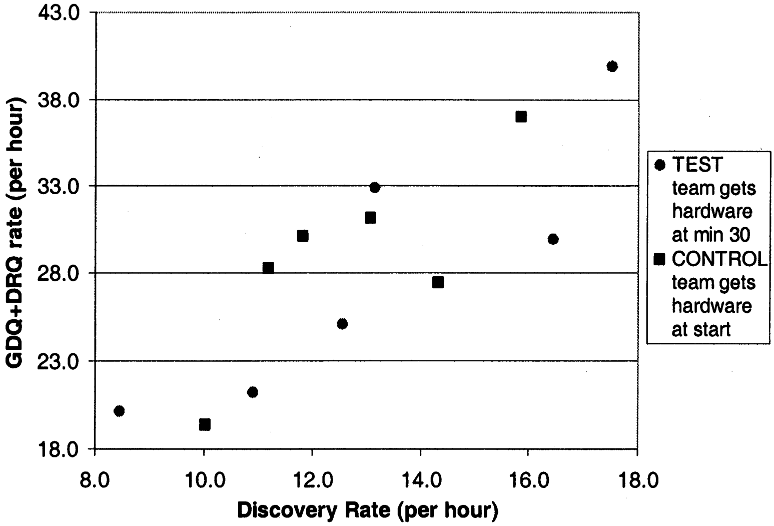


Figure 7-9. Discovery making rates of the twelve design teams plotted against their prototype score. Data points marked by squares belong to the teams in the control group. Data points marked by circles belong to the teams in the test group.

Statistical analysis of the data plotted in Figure 7-9 yielded significant correlation coefficients for both the control and test teams (Table 7-14). However, the correlation for the test group was not as strong or significant as the correlation for the control group.

Table 7-14. Correlation coefficients (adjusted R^2) and significance values for correlation between discovery making rates and prototype scores. Bold numbers indicate strong correlation or high significance. Lighter numbers indicates weaker/no correlation or lower/no significance.

	Control R^2	Test R^2	Control P	Test P
Discovery vs. Score	0.64	0.54	**0.036**	0.058

7.2.3.3 Discovery Rate and Question Asking

Even though I had not constructed a hypothesis relating discovery making and question asking, it was natural to consider if the positive correlation demonstrated in the previous section between the discovery rates and the prototype scores of the design teams was in part related to the high scoring teams asking more DRQs and GDQs.

Statistical analysis of the data on discovery making and DRQ+GDQ asking yielded strong correlation with high significance for the test group, and significant correlation for the control group (Table 7-15).

Table 7-15. Correlation coefficients (adjusted R^2) and significance values for correlation between discovery making and DRQ+GDQ asking rates. Bold numbers indicate strong correlation or high significance. Lighter numbers indicates weaker/no correlation or lower/no significance.

	Control R^2	Test R^2	Control P	Test P
Discovery vs. DRQ+GDQ	0.55	**0.71**	0.056	**0.022**

These results suggest that the positive correlation between the discovery rates and the prototype scores of the design teams were in part related to the high scoring teams asking more DRQs and GDQs. Therefore, in future research, it would be interesting to search for patterns that might reveal causal links between the instances of discovery making and occurrences of DRQs and GDQs.

7.3 Revisiting the Hypotheses

The results enabled me to evaluate the four hypotheses outlined in section 6.4. I will now revisit each hypothesis and discuss its validity in light of the findings.

In considering H1, the qualitative analysis presented in section 7.2.2 demonstrated the following:

1. Specific question asking rates and question types are associated with each design phase.
2. The fundamentals of how design teams structure their design tasks can be uncovered by monitoring the frequency, type, and content of the questions they ask while designing.

Therefore, focusing on the flow and nature of the questions asked by design teams can serve as a roadmap to their design thinking, and provides a basic understanding of their design process. This finding validates H1: Question timing and question type are descriptive characteristics of design cognition and process.

The validation of H1 establishes the necessary context for considering H2. In section 7.1.6, I reported that the trained coders did not experience any significant difficulties in coding the identified questions according to the 22 categories of the taxonomy of questions and the DRQ-GDQ distinction. Those qualitative observations contribute to demonstrating that the principles of the taxonomy of questions and the DRQ-GDQ distinction are relevant and meaningful.

Also, the statistical analysis presented in section 7.2.2.3 demonstrated a strong and significant correlation (adjusted R^2 values of 0.68 for the control group and 0.70 for the test group with $p < 0.05$) between the combined DRQ+GDQ asking rates of the design teams and their design performance, whereas a correlation could not be demonstrated between the asking rate of any single type or class of question and design performance. Further analysis presented in section 7.2.2.5 showed that DRQs and GDQs need to be treated as complementary pairs when it comes to establishing their value as a design performance metric.

When considered in conjunction, these findings validate H2: There exists two specific classes of questions, termed Deep Reasoning and Generative Design questions. Their incidence during design activity strongly correlates with design team performance and can be taken as a performance metric.

Testing H3 entailed analyzing the postulated influence of the main intervention in the experiment—delaying the introduction of the prototyping hardware to the test teams—on the question asking behavior of design teams. Statistical analysis presented in section 7.2.2.4 demonstrated that the average of the GDQ+DRQ asking rate for the test group decreased significantly, while it did not exhibit any meaningful change for the control group between parts A and B of the experiment. Further analysis showed that the decrease in the average of the combined DRQ+GDQ asking rate of the test teams was directly associated with the decrease in the average of their GDQ asking rate.

Those findings validate H3: Question asking behavior of design teams is influenced by their access to hardware. DRQ+GDQ asking rates of design teams change when they transition from working in the absence of hardware to working with hardware.

In considering H4, I tested for correlation between the discovery making and the DRQ+GDQ asking rates of design teams. The analysis presented in section 7.2.3.2 yielded significant correlation for both the control and the test teams (adjusted R^2 values of 0.64 for the control group with $p < 0.10$ and 0.54 for the test group with $p < 0.05$). However, there is a significant limitation associated with the generalization of this finding.

Since I formulated H4 in a latter stage of this research—while evaluating the pilot experiments—the framework I developed in order to characterize and operationalize the phenomenon of discovery making had not reached the necessary depth for drawing conclusions from the results by the time the above analysis was conducted.

Therefore, this finding reiterates the importance of H4, and validates it partially: There is a significant correlation between the frequency of discoveries made by design teams and design team performance. Although this finding is highly relevant and encouraging, the framework leading to the analysis needs to be developed further and the significance of the correlation needs to be higher ($p < 0.05$) for discovery making to be justified as a performance metric.

Chapter 8

SYNTHESIZING A QUESTION-CENTRIC DESIGN THINKING MODEL

A question-centric design thinking model, which describes a structure for design thinking, can be synthesized from the findings of this research. This entails reconsidering the empirical findings within the context of the theoretical frameworks on the nature of questions asked while designing and design performance. My synthesis method consists of the following steps:

1. Assigning meaning to the empirical findings by developing three paradigms that treat question asking in design as a:
   - Process
   - Creative negotiation act
   - Mechanism for managing divergent-convergent thinking modes
2. Using the third paradigm to outline a process for arriving at design decisions by asking questions.
3. Considering the implications of the verified hypotheses in light of these three paradigms.
4. Operationalizing the key elements of the insights gained in the preceding steps by mapping them onto the design process.

In the following three sections, I present the three paradigms outlined in the first step. In the fourth section, I outline the implications of the verified hypotheses. In the fifth section, I present the outcome of my synthesis, a question-centric design thinking model. In the final section, I consider five potential applications of the model.

8.1 Question Asking as a Process

Two frameworks were developed in Chapters 3 and 4. The first framework is a comprehensive taxonomy of questions asked while designing. It characterizes and differentiates questions according to their conceptual meaning (Table 3-1). The resulting taxonomy is hierarchical as the lower level question categories are associated with less sophisticated cognitive mechanisms than the higher level categories. Of particular interest were two classes of questions encompassing the higher level categories: Deep Reasoning Questions (DRQs), which reflect convergent thinking, and Generative Design Questions (GDQs), which reflect divergent thinking.

The second framework conceptualizes design performance in terms of the relationships between four phenomena: design performance, design cognition, design process, and question asking (Figure 4-6). The relationships are hierarchical as the lower level phenomena are thought to be a subset of the descriptors of the higher level phenomena. Design cognition and design process are considered to be descriptors of the same level as they are strongly dependent on each other in the sense that they feed into each other in a cyclic fashion.

The hierarchical structure of the framework on the nature of questions suggests the possibility and relevance of treating question asking as a process. However, since it only articulates the conceptual differences between questions, its principles alone are not sufficient in forming a process-centric view of inquiry in design. Although the hierarchy suggests temporal distinctions, it does not address them explicitly. However, the timing of questions, an element of inquiry investigated in the experiments, provides an initial understanding for the missing temporal dimension. Moreover, considering the empirical findings in conjunction with the principles of the hierarchy strengthens the meaning and validity of treating question asking as a process; the principles of the hierarchy can relate a process-centric view of inquiry to the design processes of teams, and ultimately, to design performance.

The rationale presented in the preceding paragraphs is an advanced formalization of what Baya and I have independently observed in the question asking behavior of designers. Baya wrote: “The questioning behavior is not random. New questions are being asked after reflecting on information received in answer to a question” [Baya 1996]. The findings of the research presented in this book not only reiterate Baya’s observation, but also build on it by formalizing several key aspects of the inquiry process in design.

Specifically, fundamental dimensions of that process can be summarized as follows: Low level questions, those that do not belong to the DRQ or GDQ classes, need to be asked in order to verify and clarify facts, identify and acquire relevant information, form the necessary communication base, and mediate social interaction. Only then can the higher level Deep Reasoning and Generative Design questions, whose function will be discussed and illustrated in section 8.5, be asked effectively. It is important to stress that the lower level questions do not have "low" value. They are qualified as being "low" simply because they need to precede the higher level questions. Attempting to ask the higher level questions without asking the lower level questions first would cause the questioner to build upon an inappropriate understanding and inevitably result in poor performance. However, asking the lower level questions and building an appropriate information and communication base does not guarantee high performance.

8.2 Question Asking as Creative Negotiation

Three significant findings on the use of Proposal/Negotiation questions by the design teams during the experiments are reported in Chapter 7:

1. Approximately 10% of all of the questions asked belonged to the Proposal/Negotiation category (the second most frequently asked question type after the Verification type).
2. Approximately 40% of all of the Deep Reasoning and the Generative Design questions belonged to the Proposal/Negotiation category (the design performance metric established in this research is the frequency of occurrence of DRQs and GDQs).
3. The Proposal/Negotiation questions were most influential during conceptualization and implementation phases of the design process.

These findings demonstrate that Proposal/Negotiation questions play a critical role in the inquiry processes and performance of design teams. However, they do not provide specific insight into the mechanism(s) through which that role is fulfilled. Qualitative consideration of the empirical data provided a level of insight by revealing one such mechanism.

During this consideration, focusing on the temporal dimension of question asking presented me with a meaningful dilemma: did the concept(s) in the question exist prior to the formulation of the question, or did the formulation of the question lead to its/their creation? These two questions proved to be instrumental in establishing a context for comparing the temporal dimensions of GDQs with DRQs. Although this dilemma cannot be truly resolved since the creation of concepts cannot be treated as a discrete phenomenon—even if it could be, there is no method to directly measure the phenomenon as it occurs in a designer's mind—I will consider it to illustrate the insight I have gained.

The concepts in DRQs exist prior to the formulation of the question. For example, the unknown concept in the Causal Antecedent question: "Why is the wheel spinning?" points to a concept associated with an event that has already taken place—the wheel spinning—and therefore, already exists. Conversely, the concepts in GDQs are created after the formulation of the question. For example, the unknown concept(s) in the Scenario Creation question: "What if the device was used on a child?" points to concept(s) associated with a hypothetical event, and therefore, will be created after the question is formulated. (A detailed discussion on each question category can be found in section 3.3.)

Proposal/Negotiation questions constitute an exception; the concept(s) in a Proposal/Negotiation question can already exist, or be created after the formulation of the question as a consequence. More importantly, they can also be created during the formulation of the question since most Proposal/Negotiation questions play a transitional role by simultaneously pointing at past and future events or states. This establishes a high degree of conceptual continuity in discourse.

In a team setting, conceptual continuity enables designers to build on each other's ideas and work more effectively as a group. For example, if the interaction building up to the question: "How about using the wheel instead of the pulley?" is considered, it is very likely that the concept "using a wheel" has occurred to the questioner right before the communication of the question while he/she was formulating the question, and that the concept "using a pulley" had been proposed earlier by another person. While asking the question, the questioner creates a spontaneous link between a proposed concept (in the past) and a newly generated hypothetical concept (in the future).

This type of cognitive interplay that Proposal/Negotiation questions promote constitutes a mechanism for influencing the design performance of teams, and supports the notion of treating question asking as "creative negotiation."

8.3 Question Asking as a Mechanism for Managing Convergent and Divergent Thinking Modes

The findings reported in section 7.2.2.2 demonstrate that design teams rely on GDQs during conceptualization, and DRQs during implementation and assessment (Table 7-7).

More specifically, during conceptualization, design teams rely on GDQs as agents of divergent thinking, which entails reframing of previously recognized needs and other existing understandings that establish context, generation of alternatives, and negotiation (and as discussed in the previous section, creative reproposal) of design concepts. These events contribute to preserving or increasing ambiguity[25]. The formulation of GDQs in order to initiate divergent thinking modes is not random. Rather, it is a conscious effort on behalf of design teams, a response to a need for conceptual expansion and creativity. Design teams continue to pose GDQs and exhibit divergent thinking until that need is satisfied.

During implementation and assessment, design teams rely on DRQs as agents of convergent thinking, which entails focusing on solutions, reiterating and focusing on goals, seeking and establishing causality, and reducing the number of alternatives. These events contribute to reducing ambiguity. As is the case with GDQs, the formulation of DRQs is not random. It is a response to a need to move toward design decisions and specifications. Design teams continue to pose DRQs and exhibit convergent thinking until that need is satisfied.

This comparison does not imply that design teams simply stop asking DRQs when exhibiting divergent thinking, and stop asking GDQs when exhibiting convergent thinking. As mentioned earlier, what I mean by a team "relying" on a specific class of questions is that a class of questions playing a comparatively more influential role in the team's progress toward meeting design goals than the other classes of questions. In many cases, that also means that the design team asks a higher number of GDQs when conceptualizing compared to the number of GDQs it asks when implementing and assessing, and vice versa, which results in the DRQ/GDQ ratio to change. The findings on DRQ+GDQ asking rates of design teams

---

[25] Ambiguity refers to the level of conceptual abstraction. For example, a car can be described as a transportation device, or as having, among other features, four wheels. The latter description is at a lower level of conceptual abstraction, and therefore, less ambiguous than the first description.

when working with and without hardware support this observation[26]; the DRQ/GDQ ratio increased due to a slight increase in the DRQ asking rates and a significant decrease in GDQ asking rates for the test teams when they transitioned from working in the absence of hardware to working with hardware.

These relationships between GDQ-DRQ usage and divergent-convergent thinking of design teams suggest and support the notion of treating question asking as a mechanism for managing divergent and convergent thinking modes.

8.4 Implications of the Verified Hypotheses

When the verified hypotheses are considered in conjunction with the discussion in the previous sections of this chapter, the following conclusions can be drawn:

1. The process of inquiry reflects key aspects of design thinking and design processes of teams. Furthermore, the design thinking of teams evolves while asking questions. While formulating questions—formulation of each question can be considered to be a micro-design task—design teams create the opportunity to structure their design thinking by diverging and converging on design concepts.
2. The frameworks developed in Chapter 3 for characterizing and differentiating questions according to their conceptual meaning, and in Chapter 4 for measuring design performance, are valid, and have potential for further development.
3. The question-based metric derived in this study not only measures design performance, but also serves as a descriptive "lens" for revealing and monitoring the thinking of designers during design activity.
4. Question asking, hence design thinking, of teams is strongly influenced by their access to hardware. When conceptualizing in the absence of hardware, design teams exhibit more divergence in their thinking by relying more on Generative Design Questions. Controlling access to hardware could provide a means to regulate the convergent and divergent thinking of design teams.

---

[26] Although the test teams went through all three design phases in the experiment when working with and without hardware, they conceptualized more when working without hardware, and implemented and assessed more when working with hardware.

8.5 A Question-centric Design Thinking Model

A question-centric design thinking model, which describes a structure for design thinking, can be synthesized by following the method outlined at the beginning of this chapter. The key elements of the insights gained in the preceding sections of this chapter are recapped in the following conclusions regarding GDQ-DRQ utilization, divergent-convergent thinking, design process, and design performance:

During conceptualization, Generative Design Questions are instrumental in *preserving* or *increasing ambiguity* by:

- Reframing previously recognized needs and understandings
- Generating alternative design concepts
- Creatively negotiating proposed design concepts

During implementation and assessment, Deep Reasoning Questions are instrumental in *reducing ambiguity* by:

- Reiterating goals
- Focusing on deliverables
- Seeking and establishing causality
- Reducing the number of proposed design concepts

High performance design teams realize the importance of managing ambiguity, and use the GDQ and DRQ instruments in a balanced fashion to operate at the necessary level of conceptual abstraction throughout the design process. Therefore, the manifestation of convergent-divergent thinking in the question asking and decision making processes of design teams in the form of Deep Reasoning and Generative Design Questions constitutes a performance dimension in design activity.

The resulting design thinking model illustrates the transformation of design requirements into design concepts through Generative Design Questions, and the transformation of those concepts into design decisions and specifications through Deep Reasoning Questions (Figure 8-1).

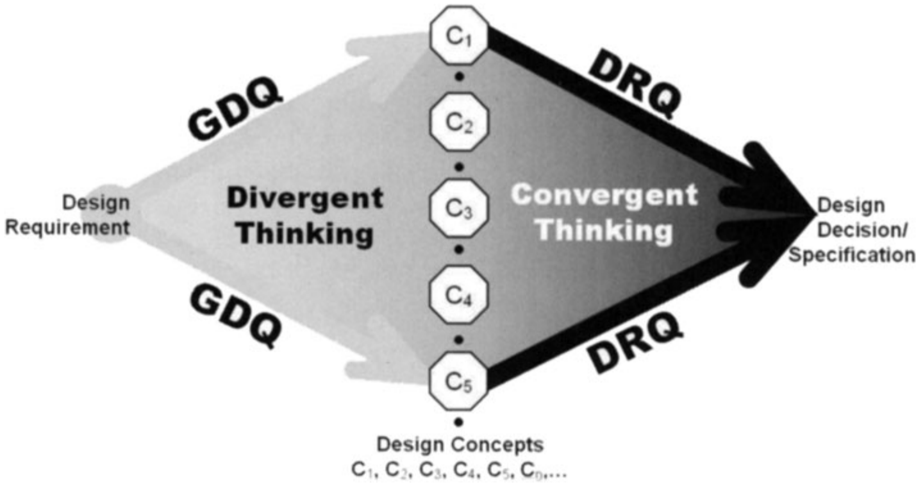


Figure 8-1. A question-centric design thinking model illustrating the transformation of requirements into design concepts through Generative Design Questions (GDQs), and the transformation of those concepts into design decisions through Deep Reasoning Questions (DRQs).

8.6 Potential Applications of the Design Thinking Model

In this section, I will discuss five potential applications of the design thinking model, and identify the principal research questions associated with them.

8.6.1 Increasing Design Performance by Promoting the Asking of more DRQs and GDQs

In the short term, a pragmatic and potentially rewarding research question to address is: Does the correlation demonstrated in section 7.2.2.3 between the combined incidence of DRQs and GDQs and design performance result from a causal relationship?

Answering this question would require the development of a method that promotes the asking of more DRQs and GDQs by design teams. The method would then need to deployed as an intervention, and its effect on design performance would need to be measured.

If the intervention results in increased performance, a strong case for a causal relationship can be made. The design thinking model would be validated, and proven to be directly applicable to design practice.

However, if the intervention does not result in increased performance, DRQs and GDQs might simply be a surrogate for other, and perhaps less visible, cognitive phenomena. In that case, the underlying cognitive phenomena would need to be identified, understood, and augmented to improve design performance.

8.6.2 A Framework for Discoveries, Questions, and Performance

Another research task that follows directly from the findings of this work is a more detailed analysis of the relationships I have outlined between asking DRQs and GDQs, design performance, and discovery making. That would entail constructing an analytical framework that characterizes and operationalizes discovery making while designing, and using that framework in order to identify potential relationships between DRQ+GDQ sequences, instances of discovery making, and design team performance.

8.6.3 Real-Time Determination and Display of the Question Asking Metric: An Instrument for Raising Design Team Performance Awareness

The question asking performance metric can be developed into an instrument that measures and displays design team performance in real-time. That instrument would provide team performance information to design teams and others who share responsibility in their success, such as coaches and managers, increasing performance awareness.

Design teams can monitor their progress with the instrument while they design. Support personnel, such as coaches, who traditionally do not have access to direct methods for evaluating the performance of the design teams they are meant to support, can utilize the instrument to obtain a real-time understanding. That would give them the ability to time and characterize their support more effectively, which often comes in the form of constructive interventions.

However, the instrument would have limited utility if it were not automated. Real-time automation can possibly be achieved in software by transcribing digitized discourse data, and analyzing the transcripts in order to identify occurrences DRQs and GDQs. However, these are non-trivial tasks, and would undoubtedly pose significant challenges.

8.6.4 Design Information and Knowledge Systems

Currently, there is a strong interest in the design research community to develop design information and knowledge capture and reuse systems. It is imperative for such systems to incorporate query based interfaces when indexing, accessing, and sharing information. The descriptive findings of this research can play a significant role in designing such interfaces, and be translated into requirements that need to be met if the systems are to support the thinking of designers effectively.

8.6.5 Toward a Unified Question-Decision Centric Theory of Design

In the long term, a significant contribution would be to integrate the findings of this research on question asking with existing knowledge on decision making in constructing a design theory. Such an approach can be structured by expanding on the two axiomatic dependencies discussed in Chapter 2 regarding questions and decisions: every question operates on decisions as premises, and conversely, every decision operates on questions as premises.

The implication is that current decision making models assume the availability of pivotal information when advocating decision making methods without addressing the mechanisms for acquiring the information, and that if those models are viewed in light of these dependencies, question asking can be taken to be one such mechanism. Developing that approach might result in a new design theory unifying decision making and question asking processes, where question asking would attain equal rank as decision making since high quality questions would yield high quality information. In other words, decision making could be viewed as taking place *during* question asking, and vice versa. Validating this concept and implementing it in the form of a software tool would have the potential to impact decision making in engineering design practice.

Appendix

A. Subject Instructions for the Test Group

Exercise Description/Product Requirements

In this exercise, you will be asked to design and prototype a "bodiometer"; a device that can be moved along the contours of male and female bodies to measure the distance traveled, and hence, the length of body segments—namely, the handweb and the head circumference. The bodiometer must be built from a LEGO parts kit which costs 30 dollars and contains a variety of structural and mechanical components, but no electrical components. No other materials or parts except those supplied with the kit are allowed. Pencil marks may be applied prior to operating the device.

Performance Criteria

Handweb is the perimeter of a hand measured from one side of the wrist to the other, including both sides of the fingers. **Head circumference** is the circumference of the skull measured at eyebrow level.

What drives the **overall team score** is a combination of sales and cost of your device. The factors that affect sales and cost are explained below. There will be 11 other design teams carrying out the same exercise. Each team's objective is to maximize their score. Scores will be computed using the following equations:

Score = Sales – Cost
Sales = Design Concept + Aesthetics + Measure Time – Error
Cost = Number of Parts + Manufacturing Time

Variables in these equations are defined as follows:

Error is scored as the cumulative absolute value (10 points for 1 inch of error) of the difference between the sum of the two team measurements and the official measurement where:

Team-measurement = Handweb + Head Circumference
Error = Absolute Value {(team measurement)-(official measurement)}

Design Concept is a bonus for a design that provides an instrumented readout, and is worth 50 points. Instrumented readout is any method which allows the user to "read off" a measurement by simply looking at the device without making any calculations or looking at any value tables.

Aesthetics is a subjective Bonus category (0-10 points), computed by averaging the scores handed out by a panel of judges (3 design researchers other than the experimenter). Opinions will be based on the device itself. Visual and "intellectual" aesthetics may enter into this opinion.

Measure Time is the combined time it takes for the judges to make the two measurements. Sales points will be earned in this way (lower time scores higher): 1st=15, 2nd=13, 3rd= 11, 4th= 10, 5th= 8, 6th= 7, 7th=5, 8th=4, 9th=3, 10th=2, 11th=1, 12th=0.

Number of Parts is the total number of parts used in your design. Cost points will be given in this way (higher number scores higher): 1st=15, 2nd=13, 3rd= 11, 4th= 10, 5th= 8, 6th= 7, 7th=5, 8th=4, 9th=3, 10th=2, 11th=1, 12th=0.

Manufacturing Time is the time it takes to rebuild the prototype from an identical and new parts kit after the main part of the experiment is over. Cost points will be given in this way (higher time scores higher): 1st=15, 2nd=13, 3rd= 11, 4th= 10, 5th= 8, 6th= 7, 7th=5, 8th=4, 9th=3, 10th=2, 11th=1, 12th=0.

SUGGESTED Schedule and Process

Phase I--90 minutes

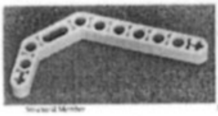
0:00-0:10: Teams receive the Project Requirements and Performance Criteria worksheet and are encouraged to ask for clarification.

0:10-1:30: Concept Generation and Prototyping: The purpose of Phase-I is to explore the design requirements, generate design concepts, and prototype one way of meeting the Product Requirements. The LEGO kit will be provided to you at the beginning of this phase. The deliverable is a functional physical prototype.

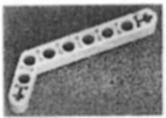
Phase II--5 minutes

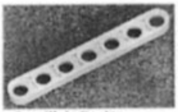
1:30-1:35: Manufacturing: In this phase, you will be asked to build a replica of your prototype from an identical and new LEGO parts kit. You may use your existing prototype from Phase II as a reference. The time it takes you to build the replica will be measured and taken as an indicator for the manufacturing time of your design.

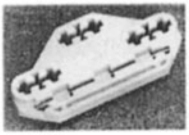
B. Prototyping Hardware Catalog for the Test Teams


Structural Member


Structural Member [illegible]


Structural Member

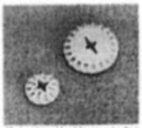

Structural Member [illegible]


Structural Member

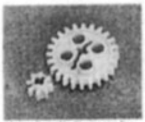

Structural Member


Pulley [illegible]

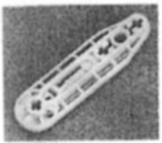

Pulley [illegible]

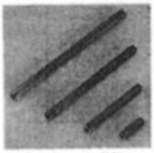

Meshing Gears [illegible]


Meshing Gears [illegible]

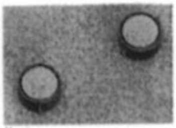

Meshing Gears [illegible]

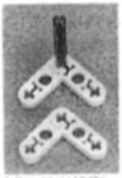

Structural Member [illegible]

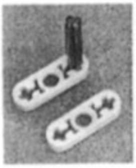

Structural Member


[illegible]

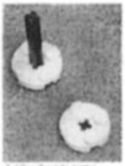

Structural Member


Magnets


[illegible]

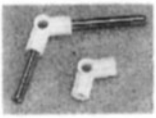

[illegible]

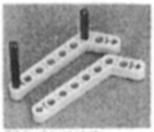

[illegible]

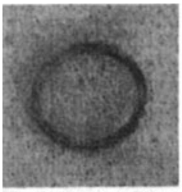

[illegible]


[illegible]

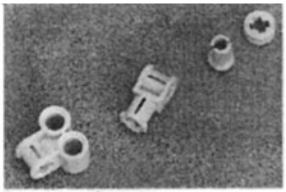

[illegible]


[illegible]

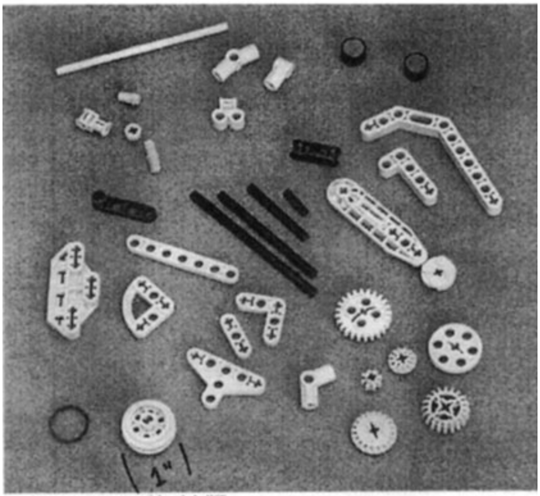

[illegible]



Rubber Band (stretches to 100+ mm)



Rod connectors and stoppers



Sample parts representative of the whole KIT

C. Sample Transcript (Design Team 1)

Time	Sub	Utterance
00:00	E	Now this is the real thing. Here's the instructions. It's two pages long. There's something on the back, too. So what I'll let you do is just let you read through it once. And during the exercise, I'll be right outside in this other room. So if you have any questions you can come and just get me. If you knock on this door I'll just come back into the room and we can ask the questions. But I'll just be here for five minutes just to make sure, once you read it, everything's clear. You can still ask questions later but, you know, I'll just be here for five to ten minutes. The schedule's on the back, but you should just kind of read through it, the way it is.
01:27	A	YY
01:31	B	Wrist. Maybe it's here, besides your fingers.
01:38	C	I wonder if it's this way?
01:41	B	Ask for it. Oh. Okay. Alright. Fingers. (pause) So it has to be really small.
02:11	A	YY
02:41	B	Are we actually trying to make this thing?
02:43	E	Yeah. You will, yeah you will prototype it with the Lego kit. Yeah.
02:46	B	Okay.
	(E brings in Lego kit)	
03:40	B	Okay. Star Wars.
03:48	C	Do we get to keep this?
03:49	E	Yeah. Yeah sure. (pause) So just for your information, I'm running this experiment in two ways. So other groups, you know there's two batches of groups and one group will do it one way and one group will do it another and then I'll compare the two. But I can't tell you before the experiment how they're different-
04:08	B	-Sure thing.-
04:09	E	-But; you know, so you might not be able to sense what's different.
04:11	B	Okay.
04:12	E	I'm just letting you know.
04:25	E	But both groups are evaluated based on the same, both types of groups will be evaluated based on the same point scheme.
06:07	B	I think the problem's going to be around the hand because you're limited by the space. If that can measure the hand accurately then we'll do okay measuring the skull [...]
06:33	C	What are we going to try to do? Maximize XX?
06:35	B	Yeah.
06:36	C	Minimize XX?
06:37	B	Yeah.
6:38	C	Do you know which, shall we try to concentrate on one of these or try to XX?
06:44	B	We should just brainstorm pulling out concepts.
06:46	C	Yeah.
06:53	B	So, are we, can we start anytime?
06:55	E	Yeah. Sure.
06:56	B	Okay. Let's look at the parts we have.
07:01	A	We could brainstorm without the parts.
07:04	C	Yeah. I think that's the best way.
07:07	A	So we, we're not limited by them.
07:09	B	Alright. Cool. Let's do that.
07:28	A	Can we use the board?
07:19	E	Yeah, you can use the board. It's on the camera. I can also bring you a sketch pad. I'll go get it.

07:27	B	Go ahead. Yeah. Why don't you give YY of designs and I'll go-
07:33	C	-Let's just talk about how we want this thing to look like. Like what it's features are going to be.
07:40	B	Well, ideally-
07:41	C	-Sort of a wheel, Right?
07:43	B	Yeah. Ideally I want a wheel.
07:45	C	It must have a wheel.
07:46	A	Why?
07:47	C	To measure with.
07:48	B	Not necessarily because if we don't have that part, if we don't have a round part.
07:52	C	Wait is it something that's going to be able to move by itself or are we going to actually move it?
07:59	B	We are going to move it.
08:00	A	We are going to move it.
08:01	B	There is no electrical parts.
08:02	A	Yeah.
08:04	B	Yeah, I was thinking it would be like a very small container with the wheel--with some sort of--
08:09	C	--we'll be counting--
08:09	B	--rubber-
08:10	C	-how many times it goes around-
08:12	B	-edge. Yeah exactly.-
08:13	C	-and calculate the circumference.
08:14	B	Exactly
08:16	A	The number of turns. (pause) I was thinking more of something like a string. Okay. Just brainstorming. I don't know how we'd do it with Lego's. You could put a string around it and then stretch it and measure it. That's going to tell you how much it-
08:38	C	-And then, how accurate is it going to be? It's not going to like stick to the hand.
08:44	A	That's true.
08:45	B	Are we allowed to use-
08:46	E	-Yeah. You can use the tape measure.
08:47	B	-use a tape? So for a measurement?
08:49	E	Yeah. And the string if you want to measure it, your head or whatever, perimeter. That's how the official measurements are going to made. By using a string and tape measure.
09:00	A	Are we going to be able to use this for, in combination with the Lego, what?
09:07	E	No. No.-
09:08	B	- It's just the Lego parts.
09:09	E	You need to use those parts. Yeah. Nothing outside of those parts.
09:12	C	Okay, so, we're basically using that to make it. Just a Lego?
09:15	B	Yeah.
09:16	E	Yes. That's right.
09:17	B	Yeah. So I don't know if we should-
09:18	A	-So we can't even do that.-
09:19	B	-yeah. I don't know if we should look at it. The parts.
09:22	A	Yeah.
09:23	B	Because we're totally limited by the parts. [spreading out Lego's] Well. We got a wheel.
09:30	A	That's too big.
09:31	B	We got.

09:33	E	I'll be right outside.
09:34	B	Okay.-
09:34	C	-Alright-
09:35	B	-Thanks.
09:42	A	YY
09:44	B	Yeah.
09:46	C	Are we? Are we being recorded?
09:49	A	Yup.
09:57	A	We could also […]
09:59	B	Umm.
10:00	A	So-
10:01	C	-Something that-
10:02	A	-wheels-
10:03	C	-that counts how many turns. Cause if the wheel's too small, are we going to be able to, like, read it off with our eyes?
10:11	A	Right.
10:13	C	Well I guess that's, that's what we have to do.
10:15	B	We don't have anything XX.
10:16	A	Right.
10:24	B	Yeah. Ideally, I mean, it would be nice if there was, like a détente, which clicks, like with every, every revolution.
10:31	A	Yeah.
10:31	B	Right?
10:32	A	See here. We're allowed to make to make a mark-
10:34	B	-a mark. Yeah. Let's make a central mark.
10:49	B	We have gears.
10:50	C	Whereas the design concept, is it bonus for a design that provides an instrumental readout? [reading] Instrumental readout is any method which allows a user to read off a measurement while simply looking at the device-
11:02	B	-Right.-
11:02	C	-without making a calculation or looking at any value tables.
11:11	B	We don't have a lot of good parts here.
11:20	B	Do you want to open this?
11:22	C	Let's open it here.
11:27	A	Not YY
11:31	B	Okay. There is a rubber seal. (pause) Rubber seals aren't good because […]
11:47	A	There are (pause) of black things.
11:50	B	Oh. It's like a belt.
12:05	C	Should we just make this? [looking at Lego plans]
	(laughter)	
12:10	B	Yeah. You should.
12:14	A	So basically we (pause) want to do this. Right?
12:21	C	Yeah.
12:22	B	Yeah. Well that's one concept. We shouldn't-
12:24	C	-That's one concept-
12:24	B	-Yeah. We shouldn't narrow ourselves down to just that. We should keep thinking what else we could measure. How else we could measure our hand. Because this is going to be the bottle neck. Right?
12:36	A	Right.
12:36	C	This is, like much harder than the skull-
12:37	C	-Yeah cause it's-

12:38	B	-Yeah. We're limited by space. (pause) Okay. We can have either the wheel. We can have a string, which is clearly not possible with this, these parts.
12:57	A	What's that?
12:58	B	Like with your string concept? You were saying that we could have a piece of string that runs around-
13:03	A	-Yeah but we can't-
13:04	B	-yeah but we can't-
13:05	C	-we can't use that one-
13:05	B	-we can't use that-
13:06	A	-No.-
13:07	B	-So what else can we do? Other than a wheel? (long pause) Well it sounds really stupid, but what about one bar that floats? Small [...]
13:29	A	YY it's almost like a string.
13:33	B	Right. (long pause) Right. A fully articulated (pause). Yeah. Basically, a mechanism which has many, many joints in very small sections. Then it is like a snake. Almost. And you can bend it around whatever profile you want.
14:07	C	It's going to be really small parts, though.
14:09	B	Yeah. It has to be really small.
14:11	C	Yeah. Because if you have things like this-
14:14	B	-Yeah-
14:14	C	-YY-
14:15	B	-it won't even. It won't even fit into your hand. Yeah. It has to be like little sections. Like these. Many of them. And they-
14:25	A	-The drawback would be that it, it's going to have a lot of parts.
14:29	B	Right. And also you won't be able to tell the measurement just by looking at it. Because you. Like after you bend it around your hand you would have to, probably like, mark it. Yeah. You would have to count, either count number of segments or you mark it, stretch it, and measure it.
14:48	C	Or you could use the same, same length parts. Then we know how much, how long one is. Like after seeing how many, how many joints we have, we have the links right away.
15:02	B	Right. Any other concepts? We want concepts. Concepts. We have gears.
15:13	A	Maybe, maybe we can make some assumptions about, like width of the fingers that we can't reach. YY although my finger's narrower than the cable.
15:25	B	Yeah. And then what would you do afterwards? You'd [...]
15:28	A	I don't know. I'm just saying shit. You might not have to measure this one here. I mean. Yeah. Otherwise, then it's going to be much simpler because this is straight, here. This is virtually straight. Straight. Straight. So we don't have to have so many joints.
15:53	C	Okay. Alright.
15:54	B	But then how do you measure (pause) your hand length at one time? Yeah.
	(laughter)	
16:04	C	Alright [Subject 1].
16:08	A	I'm just brainstorming.
16:09	B	Yeah. Yeah. I know. That's good.
16:41	B	Okay.
16:56	B	Aarrggh. Man. Come on.
17:07	A	Here [...]
	(laughter)	
17:13	A	YY volume. Then you can. For example, you have a (pause) some kind of container, filled with water.-
17:24	B	-Umm Hummm-
17:25	A	-and measure the volume by displacement. If you fill it up with water-

References

Adams, James L. *Conceptual Blockbusting: a Guide to Better Ideas*, Addison-Wesley, Reading, Massachusetts, USA, 1986.

Adams, James L. Classroom interactions during the course, “Good Products, Bad Products,” Stanford University, California, USA, 1996.

Asimov, M. *Introduction to Design*. Prentice Hall, 1962.

Baya, Vinod, et. al. “An Experimental Study of Design Information Reuse,” Proceedings of the 4th International Conference on Design Theory and Methodology, ASME, Scottsdale, Arizona, Sept. 13-16, 1992.

Baya, Vinod. *Information Handling Behavior of Engineers in Conceptual Design: Three Experiments*. Ph.D. Dissertation, Stanford University, California, USA, 1996.

Baya, Vinod, Leifer, Larry. “Understanding Information Management in Conceptual Design,” *Analyzing Design Activity*. John Wiley & Sons, West Sussex, England, p. 151-167, 1996.

Bloom, S. Benjamin: Editor. *Taxonomy of Educational Objectives, Handbook I: The Cognitive Domain*. David McKay Company, New York, USA, 1956.

Bloom, S. Benjamin: Editor. *Taxonomy of Educational Objectives, Handbook II: The Affective Domain*. Longman Group, London, England, 1964.

Bucciarelli, L. Louis. “An Ethnographic Perspective on Engineering Design,” Design Studies, Vol. 9, No. 3, 159-168, July 1988.

Bucciarelli, Louis L. *Designing Engineers*. MIT Press, Cambridge, Massachusetts, USA, 1994.

Brereton, Margot, Cannon, David, Mabogunje, Ade, Leifer, Larry. "Collaboration in Design Teams: How Social Interaction Shapes the Product," *Analyzing Design Activity*, edited by N. Cross, H. Christiaans, K. Dorst, Wiley, 319-341, 1996.

Brereton, Margot. *The Role of Hardware in Learning Engineering Fundamentals: An Empirical Study of Engineering Design and Product Analysis Activity*. Ph.D. Dissertation, Stanford University, California, USA, 1999.

Brereton, Margot. "Distributed Cognition in Design—Negotiating between Abstract and Material Representations," Proceedings of the 4th International Design Thinking Symposium, MIT, Massachusetts, USA, 1999.

Carrizossa, Ken, Eris, Ozgur, Mabogunje, Adegboyega, Milne, Andrew, Leifer, Larry. "Building the Design Observatory: a core instrument for design research," Proceedings of Design 2002, Dubrovnik, Croatia, 2002.

Cross, Nigel, Christiaans, Henri, Dorst, Kees: Editors. *Analyzing Design Activity*, John Wiley & Sons, West Sussex, England, 1996.

Costa, Jorge, et. Al. "An Analysis of Question Asking on Scientific Texts Explaining Natural Phenomena," Journal of Research in Science Teaching, V. 37(6), 602-614.

Cuff, Dana. *Negotiating Architecture: A Study of Architects and Clients in Design Practice*. Ph.D. Dissertation, University of California, Berkeley, USA, 1982.

Dieter, E. George. *Engineering Design: A Materials and Process Approach*. McGraw Hill, New York, USA, 1983.

Dillon, T. Jim. "The Classification of Research Questions," Review of Educational Research, V. 54, 327-361, 1984.

Dillon, T. Jim. "Questioning in Science," Questions and Questioning, edited by Michel Meyer, New York: De Gruyter, Chapter 4, 68-79, 1988.

Dillon, T. Jim. *Question and Teaching: A Manual of Practice*. Teacher's College Press, New York, USA, 1988.

Dorst, Kees. *Describing Design: A Comparison of Paradigms*. Ph.D. dissertation, Delft University, The Netherlands, 1997.

Eris, Ozgur, Hansen, Poul, Mabogunje, Ade, Leifer, Larry. "Toward a Pragmatic Ontology For Product Development Projects in Small Teams," Proceedings of the International Conference on Engineering Design, p. 1645-1650, Munich, Germany, 1999.

Eris, Ozgur, Leifer, Larry. "Facilitating Product Development Knowledge Acquisition: Interaction Between The Expert and The Team," International Journal of Engineering Education, Vol. 19, No. 1, 142-152, 2003.

Eodice, Micheal. *A Theory of Requirements Definition in Engineering Design.* Ph.D. Dissertation, Stanford University, California, USA, 2000.

Faste, Rolf. Classroom interactions during the course, "Ambidextrous Thinking," Stanford University, California, USA, 1995.

Flammer, A. "Towards a Theory of Question Asking," Physiological Research, Vol. 43, 407-420, 1981.

Frankenberger, Eckart, Badke-Schaub, Petra. "Information Management in Engineering Design—Emperical Results from Investigations in Industry," Proceedings of the International Conference on Engineering Design, p. 911-916, Munich, Germany, 1999.

Gedenryd, Henrik. *How Designers Work.* Ph.D. dissertation, Lund University, Lund, Sweden, 1998.

Gero, S. John, Editor. *Design Optimization.* Academic Press, Florida, USA, 1985.

Golden, L. James, Jamison, L. David. "Meyer's Theory of Problematology," Questioning Exchange, Vol. 2, No. 2, 149-163, 1988.

Graesser, Arthur, Lang, Kathy, Horgan, Dianne. "A Taxonomy for Question Generation," Questioning Exchange, Vol. 2, No. 1, 3-15, 1988.

Graesser, Arthur, Golding, Jonathan. "Questioning in Cognitive Psychology and Artificial Intelligence," Questioning Exchange, Vol. 2, No. 3, 315-324, 1988.

Graesser, Arthur, Person, Natalie, Huber, John. "Mechanisms that Generate Questions." *Questions and Information Systems*, edited by Thomas W. Lauer, Eileen Peacock, Arthur C. Graesser. Hillsdale, N.J.: L. Erlbaum, Chapter 9, p. 167-187, 1992.

Graesser, Arthur, McMahen, Cathy. "Anomalous Information Triggers Questions When Adults Solve Quantitative Problems and Comprehend Stories," Journal of Educational Psychology, Vol. 85, No. 1, 136-151, 1993.

Graesser, Arthur, Person, Natalie. "Question Asking During Tutoring," American Educational Research Journal, Vol. 31, No. 1, 104-137, 1994.

Hales, Crispin. *Analysis of the Engineering Design Process in an Industrial Context.* Ph.D. dissertation, Cambridge University, Cambridge, UK, 1987.

Hansen, Poul, Mabogunje, Ade, Ozgur, Eris, Leifer, Larry. "The Product Development Process Ontology: Creating a Learning Research Community," Proceedings of the International Conference on Engineering Design, Glasgow, UK, 2001.

Hazelrigg, G. A. "An Axiomatic Framework for Engineering Design," Journal of Mechanical Design, Vol. 121, p. 342-347, September 1999.

Hazelrigg, G. A. "On The Role and Use of Mathematical Models in Engineering Design," Journal of Mechanical Design, Vol. 121, 336-341, September 1999.

Howard, Ronald A. "Decision Analysis: Practice and Promise," Management Science, Vol. 34, No. 6, p. 679-695, June, 1988.

Hubka, V. *Principles of Engineering Design.* Translated by W. E. Eder, Butterworth Scientific, Guildford, 1982.

Ijiri, Yuri. *The Foundations of Accounting Measurement.* V. Prentice-Hall Inc., Englewood Cliffs, New Jersey, USA, 1967.

Kerry, T. "Classroom Questions in England," Questioning Exchange, Vol. 1, p. 32-33, 1987.

Kuffner, Tom. *Mechanical Design History Content: the Information Requests of Design Engineers.* Master's thesis, Oregon State University, Oregon, USA, 1990.

Kuffner, Tom, Ullman, David. "The Information Requests of Mechanical Engineers," Design Studies, Vol. 12, No. 1, 42-50, January 1991.

Leifer, Larry. Classroom interactions during the course, "Mechatronic Systems Design," Stanford University, California, USA, 1994.

Lehnert, G. Wendy. *The Process of Question Answering.* Lawrence Erlbaum Associates, Hillsdale, New Jersey, 1978.

Lindemann, Udo. "A Model Design Process of Individual Designers," Proceedings of the 12th International Conference on Engineering Design, Munich, Germany, 1999.

Mabogunje, Adegboyega. *Measuring Conceptual Design Performance in Mechanical Engineering: A Question Based Approach.* Ph.D. dissertation, Stanford University, California, USA, 1997.

Marsh, J. R., Wallace, K. M. "Integrity of Design Information," Proceedings of the International Conference on Engineering Design, p. 1449-1454, Prague, 1995.

Marsh, J. R. *The Capture and Structure of Design Experience.* Ph.D. dissertation, Cambridge University, Cambridge, UK, 1997.

McCracken, James Richard. *Questions: Assessing the Structure of Knowledge and the Use of Information in Design Problem Solving.* Ph.D. Dissertation, Ohio State University, Ohio, USA, 1990.

McGown, A. et. al. "Using Concept Sketches to Track Design Progress," Proceedings of the 4th International Design Thinking Research Symposium, MIT, Massachusetts, USA, 1999.

McMahon, C. A., Lowe, A., Culley, S. J. "An Information-Connection Model for Design," Proceedings of the International Conference on Engineering Design, p. 1651-1656, Munich, Germany, 1999.

Mendenhall, William, Beaver, Robert. *Introduction to Probability and Statistics*. Duxbury Press, Ninth Edition, California, USA, 1994.

Minneman, Scott. *The Social Construction of a Technical Reality*. Ph.D. dissertation, Stanford University, California, USA, 1991.

Miyake, N. Norman D. A. "To ask a question, one must know enough about what is not known," Journal of Verbal Learning and Verbal Behavior, Vol. 18, 357-364, 1979.

Newell, A. and Simon, H. A. *Human Problem Solving*. Prentice-Hall Inc., Englewood Cliffs, New Jersey, USA, 1972.

Otto, N. Kevin, and Wood, L. Kristin. *Product Design: Techniques in Reverse Engineering and New Product Development*. Prentice-Hall Inc., Upper Saddle River, New Jersey, USA, 2001.

Pahl, G. "How and Why Collaboration with Cognitive Psychologists Began," Designers—the Key to Successful Product Development, Darmstad Symposium, 1997.

Pahl, G. and Beitz, W. *Engineering Design: A Systematic Approach*. The Design Council, London, England, 1988.

Pugh, Stuart. *Total Design: Integrated Methods for Successful Product Engineering*. Addison-Wesley Publishing Company, Wokingham, England, 1990.

Pugh, Stuart. *"Concept Selection—A Method that Works," Creating Innovative Products using Total Design*. Addison-Wesley, Chapter 14, 167-176, 1996.

Rabinowitz, Mitchell: Editor. *Cognitive Science Foundations of Instruction*. Lawrence Erlbaum Associates, New Jersey, USA, 1993.

Radford, Antony, Gero, S. John. *Design Optimization*. Gero, S. John, Editor, p.229-258, Academic Press, Florida, USA, 1985.

Roth, Bernard. Classroom interactions during the course, "Designer in Society," Stanford University, California, USA, 1995.

Rowe, G. Peter. *Design Thinking*. MIT Press, Massachusetts, USA, 1987.

Shank, R. C. Conceptual Dependency: "A Theory of Natural Language Understanding," Cognitive Psychology, Vol. 3, No. 4, p. 552-631, 1972.

Schon, Donald A. *The Reflective Practitioner: How Professionals Think in Action*. Basic Books, New York, USA, 1983.

Schon, Donald A. "Problems, Frames and Perspectives on Designing," Design Studies, Vol. 5 No. 3, p. 132-136, 1984.

Schon, Donald A. "Teaching and Learning as a Design Transaction," Research in Design Thinking, Proceedings of Industrial Design Engineering, p. 21-34, Delft University of Technology, The Netherlands, May 29-31, 21-35, 1991.

Sigel, E. Irving. "Problem Finding in Creativity—A Review of The Creative Vision: A Longitudinal Study of Problem Finding in Art," Questioning Exchange, Vol. 2, No. 2, 141-147, 1988.

Simon, H. A. *The Science of the Artificial.* 2nd Edition, MIT Press, Cambridge, Massachusetts, USA, 1981.

Tang, C. John. *Toward an Understanding of the Use of Shared Workspaces by Design Teams.* Ph.D. dissertation, Stanford University, California, USA, 1989.

Tang, C. John, Leifer, J. Larry. An Observational Methodology for Studying Group Design Activity, Research in Engineering Design, Vol. 2, p. 209-219, 1991.

Ullman, G. David. *The Mechanical Design Process.* McGraw-Hill, Inc., 1992.

Ullman, G. David. "Toward the ideal mechanical engineering design support system," Research in Engineering Design, Vol. 13, p. 55-64, 2002.

Wilde, D.J. "Using student preferences to guide design team composition," Proceedings of DETC '97, Sacramento, 1997.

Willem, A. Raymond. "Design-Science Interactions," Design Engineering Division publication, ASME, V. 27, 323-235, 1990.

Willem, A. Raymond. "Design and Science," Design Studies, Vol. 11, No. 1, p. 42-47, 1990.

Wood, Bill. "A Methodology for Transforming Information into Design Knowledge," Proceedings of the International Conference on Engineering Design, p. 131-136, Munich, Germany, 1999.

Yang, Maria. *Retrieval of Informal Information from Design: A Thesaurus Based Approach.* Ph.D. dissertation, Stanford University, California, USA, 2000.

Yen, Samuel. *Capturing Multimodal Design Activities in Support of Information Retrieval and Process Analysis.* Ph.D. dissertation, Stanford University, California, USA, 2000.

Index