Empirical Comparisons of System Analysis Modeling Techniques

by

Andrew C. Gemino

B.A. , Simon Fraser University, 1986
M.A., Simon Fraser University, 1989
M.B.A., Simon Fraser University, 1992

A THESIS SUBMITTED IN PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR THE DEGREE OF

DOCTOR OF PHILOSOPHY

In

THE FACULTY OF GRADUATE STUDIES

The Faculty of Commerce and Business Administration
Management Information Systems Division

We accept this thesis as conforming to the required standard

The University of British Columbia

June 1999

Faculty ~~Department~~ of Commerce and Business Administration

The University of British Columbia
Vancouver, Canada

Date June 1, 1999

Abstract

The development of information systems consumes an increasing share of economic resources. Over a trillion dollars worldwide is invested in information technology annually, and this investment is growing over $100 billion a year. This investment occurs despite failure rates for large information system development projects that are estimated as high as 75%. The large investment and high failure rates combine to create the potential for significant impact from information system development practices that are able to address these failure rates.

Researchers, over the past thirty years, have studied factors that drive these high failure rates. One of the factors repeatedly mentioned in practitioner surveys is the importance of accurate communication in the "upstream" analysis and planning stage of a project. System development professionals are aided in their upstream planning through the use of information system development methods (ISDM's). ISDM's are modeling tools and techniques that are capable of representing information about an information system. Many alternative system analysis modeling techniques have been developed, yet few empirical comparisons of the alternative techniques have been completed. The lack of comparative empirical data has contributed to a proliferation of modeling methods and increased the confusion surrounding the adoption of system analysis methods by system development professionals and teachers.

This study addresses the issue of empirical comparison of system analysis modeling techniques. A new instrument and empirical method is proposed for developing a comparison of the level of "understanding" that a participant is able to create by viewing a description of a particular domain. The level of "understanding" is addressed using three measures: comprehension, problem solving, and text reconstruction. The new measures of "problem solving", suggested by Mayer in the field of Education Psychology, and "text reconstruction" or "Cloze", suggested by Taylor in the field of Communications, extend empirical instruments previously used by system analysis researchers.

To test the efficacy of the proposed instrument and method, two empirical studies were developed in this thesis. The first study used the new instrument to compare three development methods "grammars: Text descriptions; Structured Analysis (using Data Flow Diagrams and Entity Relationship Diagrams); and Object Oriented Diagrams. The study was labeled an "Intergrammar" comparison, as three grammars representing three fundamental approaches to developing an analysis model were compared.

Two propositions, in regards to the intergrammar study, were tested. The first suggested that viewing descriptions created with diagrams would lead to a higher level of understanding than viewing a description based solely on text. This hypothesis was confirmed. The second hypothesis suggested that viewing a domain description created using an object oriented grammar would lead to a higher level of "understanding" than viewing a description created using the "Structured Analysis" approach. The results confirmed the hypothesis that the group of participants using the Object-Oriented grammar scored higher in "understanding" than participants using the Structured Analysis grammar.

A follow-up protocol analysis was undertaken to illuminate why the participants using object methods scored. The analysis of these protocols indicated two things. First, participants using Structured Analysis made little use of the Entity Relationship Diagram (ERD). Second, participants seemed to favor the "object" concept when answering questions. These findings provide some empirical evidence that objects may be more "natural" cognitive constructs than those used in Structured Analysis.

The second study revisited a study Bodart and Weber's study regarding alternative grammars for the Entity Relationship Diagram. A grammar using mandatory attributes and relationships with sub types, the other using optional attributes and relationships, were compared. The grammars shared a common primary grammar, therefore, the second study was labeled an "Intragrammar" comparison. The new instrument was again used in this study.

The ontological constructs proposed in the Bunge-Wand-Weber (BWW) model were used to suggest the theoretical advantage of the grammar using mandatory attributes and relationships with subtypes. The results supported the theoretical advantage associated with mandatory attributes and relationships with subtypes. This intragammar study provided further evidence of the utility of the empirical instrument proposed in this thesis.

This study has implications for future empirical research in system analysis. The empirical instrument described in this thesis extends previous empirical research instruments with the introduction of the problem solving and the Cloze task. In two studies, the new instrument has displayed the sensitivity to differentiate between treatment groups. The results from the two empirical studies suggest that object-oriented analysis may hold advantages over traditional structured analysis, and that mandatory attributes and relationships may be preferred to optional attributes and relationships in the entity relationship grammar.

Table of Contents

List of Figures

List of Tables

Acknowledgements

A thesis is a large task, and with any large task there are always a lot of people to thank. First on my list is my wife, Kathy, who really deserves most of the credit for this weighty document. Her love, faith, patience, and encouragement kept me going when times were hard. Thanks for sharing the load little one.

Next on the list is my senior supervisor Yair Wand who treated me with respect, and who displayed a steadfast confidence in my research and teaching ability. As usual Yair, thank you for your time. I am also grateful for the time and energy that Izak Benbasat and Seong-Soo Lee were able to provide in guiding my research. Seong–Soo not only provided statistically significant help in guiding me through the general linear model, he also taught me much about the process of scientific discovery. Izak Benbasat provided the model for my empirical research and convinced me of the importance of validity.

There's a whole host of other people to thank as well. The CHAPS (Mike Brydon, Paul Chwelos, and Errol Smythe) for continuous support, numerous beers, interesting conversation, and of course my introduction to the great game of golf; my office buddy Genevieve Bassellier for her companionship and conversation while I waited at Yair's door; Blaize Reich who was partially responsible for getting me into the program at UBC; Nanda for the coding of the problem solving questions; the world famous GSS Champion Commerce Cubs softball team who provided a forum for a repressed sports reporter; Pat Shanahan who graciously offered to act as my thesis defence manager; Larry Pinfield for listening to me when no one else could; and Kamal Masri for understanding what I had to do and letting me do it. I also extend my heartfelt thanks, of course, to my rather large and loving family who are still trying to figure out why I needed to get another degree.

Most importantly, I would like to thank my father, who passed away just before I reached the finish line. I know he would have been proud, and no one is more responsible for helping me make my decision to wear the gown rather than work downtown. And finally, a special thanks to my little baby who sat calmly through my defence (albeit inside Mom's tummy). It has been a good trip and I hope the best is yet to come.

Chapter 1: Introduction

1.1 The System Development Problem

Information systems are an integral part of our lives. From everyday items such as electric shavers (200 lines of code), and Sony Walkmans (2,200 lines of code), to more sophisticated systems such as automobiles (30,000 lines of code), and the space shuttle (420,000 lines of code), information technology increasingly impacts the way we work, play and communicate[1]. Sophisticated information systems (IS) combine people, technology and software to bring services such as virtual banking centres, airline reservation systems, automated inventory control, digital telephone networks, and electronic commerce on the Internet. Information technology, as a consequence, takes up an increasing share of economic resources with a world market of over one trillion that is growing at an annual rate of approximately $100 billion per year (Mitchell et. al., p. 20, 1998).

The proliferation of software and information systems, to some observers, suggests that system development projects are often successful, and that system development methods have matured enough to provide relatively low risk and high return business opportunities. According to industry studies, however, failures in software development projects are common in many organizations. Recent examples of failures include: the $193 million Denver Airport baggage-handling system that delayed the airport opening for 18 months at a cost of $1 million a day; the $163 million loss absorbed by American Airlines in their failure to integrate Marriot Hotels and Budget Rent-a-Car into American's flight reservation system; the cancellation of the six year effort to upgrade vehicle registration at the California Department of Motor Vehicles, which wrote off the entire project after investing $45 million; and the reported $1 billion cost overrun faced by the Federal Aviation Administration's Advanced Automation System for air-traffic control.

Data from recent studies provide evidence that these large failures should not be considered special cases in application development. The Standish Group, in their survey of 343 companies, reported a bleak picture for software development projects reporting that only 16% of all software projects were delivered on time and on budget ("Chaos", 1995). The report also indicated that on average only 60% of the originally specified features were available in

---

[1] Estimates for lines of code from Ramstad, 1995.

the end product. Average time overruns were estimated at 222%, while average costs overruns were placed at 189%. These figures were echoed in a recent report on a survey (Gabig, 1996) conducted by the U.S. Department of Defence (DoD). The report noted that more than one third of the DoD's large-scale software development contracts are canceled before they are completed, and that the average time overrun for their projects is 150%.

It's not just information technology groups sounding the alarm. An article entitled "The Software Crisis" (Gibbs, 1994 p.87), in Scientific American, noted:

> *"The average software development project overshoots its schedule by half; larger projects generally do worse. And some three quarters of all large systems are operating failures that either do not function as intended or are not used at all."*

A report from Canada's Auditor General (October 5, 1995) provided a long list of software project failures and delays experienced by the federal government. The report noted that "*In recent years, the amount spent on developing large-scale systems, in both the private and public sectors, has risen dramatically,*" and that "*Historically, many of these systems have failed to meet the needs of prospective end-users.*"

Trends in Success Rates: It's getting better all the time. Isn't it?

Recent decades have seen remarkable advances in computing technology. It is, perhaps, a commonly held belief that advances in computing methods and technology are necessarily linked to a higher success rate for software development projects. Arguably, this has not been the case. Technological advances have had a varying degree of impact on computing in business, but these advances have not been able to stem the tide of system development failures. As evidence for this claim, a particularly disturbing statistic from the aforementioned Standish Group report noted that over 48% of executive managers surveyed indicated that they perceived more development failure now than five and even ten years ago.

Individuals inside the software development industry have recognized their industry's poor performance in system development. As evidence, Keith Brown, past president of the Canadian Information Processing Society (CIPS), the largest collection of Canadian professionals in the systems development field, noted (Computing Canada, Feb. 15, 1996, p. 1):

"Go (talk to) CEO's with significant IT departments in banks, insurance companies, utilities, and department stores.... go and ask them if they don't have concerns about the IT department's failure to deliver quality systems. Every one (CEO) will tell you they are worried... They think our industry is a bunch of shysters."

Unfortunately, these failure rates and the attention they attract are likely to persist. The increasing complexity associated with newly developed systems coincides with a significant increase in the level of dependency that many organizations place on information technology. As an example of financial dependency, a survey by Stratus Computer Inc. ("Survey", 1992) of 450 IS executives from Fortune 1000 companies provided a combined total of $3.4 billion annually spent in unplanned computer downtime due to software failure with an average hourly revenue loss of $74,191. The increased dependency was directly noted by 75% of the executives in the survey who indicated that their companies were more dependent on their information systems than they were in previous years. Together, increasing complexity and dependency imply that successful implementation of system development projects will become both more difficult and more important in the future.

Why do Information System Development Projects Fail?

Information system development is a difficult process. Like other significant business development projects, each system development project must weather an array of technological, political, and resource constraints to become successful. Many researchers accept that a percentage of software development projects will fail. The low success rates currently experienced in system development, however, are clearly unacceptable to both practitioners and researchers. For this reason, researchers over the past thirty years have studied factors that drive the high failure rate.

This research has identified a variety of factors affecting project success. A small sample of these factors include: the need for early process management and risk assessment (McFarlan, 1981); the inclusion of users in the design process (Lucas, 1975); a poor understanding of the politics in design and development (Robey & Marcus, 1984); the tendency to underestimate the cost or scope associated with the project (Kemerer, 1987); the lack of structured development techniques (methodology) for developing a system design (Coad & Yourdon, 1991); and the effect of social and organizational forces on the change process (Hirschheim, 1996).

One of the most common factors repeatedly attributed to system failure, as noted in surveys of practitioners, is the importance of accurate communication early in the system development process (Boar, 1982; Crockett, 1989; Hutching & Knox, 1995; Holdblatz & Beyer, 1995; Chaos, 1995). These surveys help to highlight the historic difficulty of bridging the communication gap between domain experts, who understand business, and system experts, who understand technology. Methods for bridging this knowledge gap are of primary interest in the field of information systems analysis. The Standish Group Report ("Chaos", 1995) provides a useful list of factors that practitioners view as important in contributing to system development failure:

Table 1: Factors Affecting System Development Failure ("Chaos", 1995)

Rank	Project Failure Factor	% Responses
1	Lack of User Input	12.8
2	Incomplete Requirements / Specifications	12.3
3	Changing Requirements and Specifications	11.8
4	Lack of Executive support	7.5
5	Technological incompetence	7.0
6	Lack of Resources	6.4
7	Unrealistic expectations	5.9
8	Unclear Objectives	5.3
9	Unrealistic Time Frames	4.3
10	New Technology	3.7

Note that six of the top ten factors in Table 1 (1, 2, 3, 7, 8, 9) are concerned with communication in the early, or "upstream", portion of the system development process. This early phase of development is commonly referred to as **"systems analysis."** The survey results suggest that when considering factors for project failure, practitioners have recognized the importance of communication in upstream analysis and planning. The results in Table 1 and surveys noted earlier indicate that current methods used to plan and to communicate planning information between developers and users may not be performing well.

The surveys noted above represent only a small portion of the research into factors important in successful system development. This brief discussion should emphasize that system development failure can result from a combination of many factors; both ***upstream*** in the process of planning and analysis, and ***downstream*** in the development and implementation process. No "silver bullet" will solve all, or even most, of the challenges faced in system development (Brooks, 1987). Recognizing that there is no "cure-all", however, does not suggest that the system development process cannot be significantly improved. On the contrary, the myriad of problems suggests that significant potential for gains in productivity and increases in project success rates are possible. Further research into the system development process should help to secure some of these gains.

1.2 The Research Scope

The area of information system development provides a wide scope for potential research topics. The first challenge in addressing this large problem is to identify a research scope that is narrow enough to be effectively researched, yet broad enough that it touches an issue of significant potential impact. This section outlines the choices made in this thesis in order to develop an arguably useful research scope.

Both upstream and downstream factors, as noted earlier, affect software development success. Research in both of these areas is necessary for improving software development success rates. Limited resources, however, require researchers to focus on specific areas for improvement. The research reported in this study will focus on the "upstream" activities in software development. These upstream activities include project planning and system analysis.

Two reasons are provided for focusing on upstream planning activities. First, significant leverage can be associated with solving problems early in the planning process of any project. As an example of this leverage in software projects, Boehm (1976, p.18) estimated that:

> *"...the relative cost of fixing problems during final testing and operations is 50-100 times greater than for problems detected during requirements specification."*

De Marco (1979, p. 9) echoes Boehm's estimates:

> *"Projects go wrong at many points. The fact that we spend so much time, energy, and money on maintenance is an indication of our failures as designers; the fact that we spend so much on debugging is an indictment of our module design and coding and testing methods. But analysis failures fall into an entirely different class. When analysis goes wrong, we don't just spend more money and come up with a desired result – we spend much more money, and often don't come up with any result."*

The opinions noted above suggest that more effective "upstream" analysis and planning will tend to reduce the number of problems early in the design process where benefits to a project can be substantial.

A second reason for focusing on analysis and planning activities is provided by practitioner surveys such as the one shown earlier in Table 1. Practitioner surveys have suggested that upstream activities - such as establishing requirements, gathering user input, and setting realistic objectives and time frames – are important factors contributing to a project's success. The motivation for focusing attention on upstream analysis and planning is based, therefore, on the leverage created by solving problems early in the development process. The quality of planning and analysis is widely recognized as an important practical factor in project success.

System Analysis Methods

One of the ways in which system development professionals can be aided in their upstream planning and analysis of software projects is through the use of ***information system development methods (ISDM).*** Examples of these ISDM's include: Information Engineering Methodology; Soft Systems Development; and Structured Systems Analysis Method (Avison & Fitzgerald, 1995). An ISDM generally outlines a process for system development and suggests one or more ***techniques*** that can be used to capture and convey relevant information about a business domain. For example, a development ***method*** might suggest that development teams begin the development process by identifying the goals of the system, then analyze the system currently in place, followed by a listing of alternatives to the current system. The same development method might also provide a set of techniques for creating diagrams or tables documenting the current or proposed systems. In this way, the system development method structures an organization's approach to building an information system.

System development methods are often used to create a set of requirements for a project, and often provide a means for communicating these requirements to system developers through the use of a graphical model (Olle et. al., 1992; Avison & Fitzgerald, 1995). The process of developing a detailed list of specifications for a system is commonly referred to as "Requirements Engineering" (Avison & Fitzgerald, 1995). The models created for requirements engineering are aimed towards developers and programmers who are focused on system design. Since these models are close to design models, requirements engineering is a process that happens somewhat downstream in the development process, and closer to implementation and coding, than system analysis. While requirements engineering is an important part of the development process, it is not the focus of this study.

Another modeling process that occurs at a higher level of analysis than requirements engineering is referred to as "Conceptual Modeling" (Mylopolous, 1992). The models created during conceptual modeling are developed in order to communicate an overview of a domain (an organizational context) to a person viewing the model. The details regarding the implementation of the system under consideration are not apparent in the model. Conceptual models are created early in the planning process (upstream) in order to develop a high level understanding of the organization being modeled. Examples of some of the graphical models that can be used for enterprise modeling include flow charts, context level data flow diagrams, conceptual entity relationship diagrams, state transition charts, object-oriented diagrams, rich pictures, and many others. Each of the diagramming techniques provides a means for communicating information about a domain to interested individuals.

The communication of graphical models is interesting from a researcher's perspective because of the different perspectives from which these models can be viewed. For example, system experts (analysts and developers) are usually well trained in development methods and graphical modeling techniques, whereas, domain experts (users of the information system) are not. Domain experts, in a similar fashion, are normally much more knowledgeable than system developers regarding work that the users must perform. These two sets of knowledge – knowledge of the domain and knowledge of the models that describe a domain – often contribute to a communication gap forming between developers and users of information systems. System development techniques have been developed as a method for potentially bridging some this knowledge gap. If the bridging of this knowledge gap between developers and users is important, then studying the effectiveness of **alternative** system analysis techniques in communicating relevant system information is a topic worthy of research attention.

The discussion above has outlined the scope of the research. The studies described in the remainder of the thesis will focus on the comparison of alternative system analysis methods that are used in the upstream phases of a system development project. Having defined the scope of the study it is necessary to choose how the comparisons between methods will be made. The next section outlines alternative approaches to the comparisons.

Theoretical and Empirical Comparisons

Two approaches can be taken in the comparison of analysis techniques: a theoretical approach and an empirical approach. The theoretical approach attempts to compare methods based a method's **theoretical** ability to **represent** different dimensions of a business domain. As an example of this approach, the Meta Model Hierarchy suggested in Oei et. al. (1992), describes a comparison based on representational power. In this approach, analysis methods are placed higher in a modeling hierarchy if they are able to describe all of the relevant features (plus at least one additional feature) provided by methods lower in the hierarchy. A method that is able to describe more of the relevant features of a business domain is considered a more general (and more preferred method) to methods lower in the hierarchy that describe less of the relevant features of a business domain.

Another ***theoretical approach*** is based on the Wand & Weber (1993) ontological[2] analysis of system analysis techniques. Examples of this approach are provided in Weber & Zhang (1991) and Green (1996). In this approach, the constructs and rules that define an analysis technique are evaluated against a set of ontological constructs. Wand & Weber propose one such set. The more complete and clear the mapping between the ontological constructs and the modeling constructs provided by a system analysis technique, the better the representational power of the modeling technique.

Proponents of a modeling technique can use the theoretical advantages to suggest that a modeling technique or analysis method should be adopted. For example, Coad and Yourdon (1991) suggest that practitioners use object-oriented analysis (OOA) methods because *"OOA organizes analysis and specification using the methods of organization that pervade people's thinking"*. This may be a theoretical advantage, but the claim is not based on any empirical evidence. Theoretical advantages may exist for many techniques in comparison with alternative techniques. Important considerations in these theoretical comparisons are the sources of the representational advantages and whether these advantages can significantly affect the communication within the analysis process. While theoretical claims provide justification for the possible sources of advantages, empirical studies can produce evidence indicating whether these theoretical advantages are evident and important in actual communication.

The focus on empirical methods in this thesis indicates a belief, held by this researcher, that the most meaningful models of domains are not written on paper, nor displayed on a screen or stored in a computer, but rather are perceived, developed, and stored in the cognition of various stakeholders in the development project. For a modeling technique to communicate effectively, it must relay and organize information that can have an effect on an

[2] Ontology is the study of how persons describe the world around them. A more complete description of Wand & Weber's approach is provided later in Chapter 3.

individual's cognitive model of a system. Arguments regarding the theoretical ability of models to capture relevant system information are secondary, therefore, to empirical questions of how these models actually might affect human cognition. Empirical studies are necessary to support theoretical idea, yet there is very little empirical evidence created by researchers. Without empirical tests, the theoretical advantages go untested and unchallenged. In this environment, it is difficult to see clear advantages in any of the analysis methods. Theoretical advantages, without associated empirical studies of these effects at the level of the individual, do not provide sufficient information for those choosing to adopt a particular modeling method. ***Empirical comparisons, therefore, are a critical component in the development of analysis methods and are the focus of this research.***

1.3 Motivation for the Research

Research regarding the effectiveness of alternative system analysis techniques has been made increasingly difficult due to the growing number of alternative techniques available to practitioners (Chatzoglou & Macaulay, 1996). Underlying each technique is the notion of a "grammar". A grammar is simply a set of constructs (symbols) and associated rules for combining these constructs. In other words, a grammar defines how a description of a domain is built. The English language has a grammar, and so do all system analysis techniques. The long list of alternative grammars underlying techniques has been called the "Methodology Jungle" by Avison & Fitzgerald (1995), or the "YAMA Syndrome" short for "Yet Another Modeling Approach" by Oei et.al. (1992). New techniques and methods continue to be created, yet few evaluations of the performance of these new methods and their related techniques are provided. This proliferation is noted by Green (1996, p. 25) who states:

> *"In the absence of sound theory for system analysis and design grammars, methodologies, techniques, and components within techniques continue to proliferate. Researchers and practitioners alike have no objective basis on which to evaluate these grammars. At the present time, one set of factors, features, or facets is as justifiable as another for use."*

This work argues for the empirical comparison of analysis techniques to reduce the tendency towards method proliferation and to improve the performance and selection of analysis techniques. It is argued that rather than concentrating effort on producing a wider variety of analysis techniques, researchers should attempt to understand how and why models are useful in the analysis process. This will help to more completely define the roles that models play in information system development, and in the end, help to produce more effective analysis techniques. These empirical comparisons will also enable researchers to gain an improved understanding of the strengths and weaknesses of analysis methods.

1.4 Research Objectives

This research, as noted above, is motivated by the high failure rate in information system development projects. Many factors, both upstream and downstream in the development process, contribute to these high failure rates. Table 1 above indicates that an important factor in the success of a project is the analysis of the domain for the proposed system. As noted, many alternative techniques exist for this analysis, yet little theoretical or empirical research is available on alternative model's comparative performance. The inability to compare methods promotes the proliferation of techniques and increases the difficulty for both practitioners and educators in choosing appropriate development methods. This study addresses the empirical comparison of analysis techniques.

Given the above motivation, this study has two objectives. The first is to suggest an empirical procedure for comparing the grammars underlying systems analysis modeling techniques. In this study, a system analysis modeling technique is defined as a set of well-defined modeling constructs along with rules for combining these constructs. The constructs and rules together will be referred to as a ***grammar.***

The second objective is to apply the suggested empirical method to compare actual grammars. Two studies will be undertaken. The first study will draw a comparison between text description (TXT), Object Oriented Analysis (OOA) and a combination of Data Flow Diagramming (DFD) and Entity Relationship Diagramming (ERD) techniques. Reasons for these comparisons will be provided in later chapters. The second study uses this new comparison procedure to revisit a study previously reported by Bodart & Weber (1996) which focuses on different forms of the grammar underlying the Entity Relationship Diagram (ERD).

1.5 Research Method

To accomplish these two objectives, a research instrument for the comparison of analysis techniques was developed and two comparative studies were designed and implemented. The empirical instrument for use in lab experiments was developed from a combination of comprehension tests, problem solving tasks – as developed by Mayer (1989) - and text reconstruction (or the Cloze tests) as developed by Taylor (1953). Chapter 3 discusses the development of the research instrument, Chapter 4 outlines the propositions relating to each of the comparisons, and Chapter 5 details the method used in the study.

1.6 Research Contributions

This chapter has suggested the need for empirical comparisons of modeling techniques. Research in this thesis can contribute to the field of system analysis in four possible ways:

1. The empirical research method proposed in this thesis may provide a new comparative tool useful for developing comparative information on system analysis methods. Combined with other evaluative methods, this method can help in establishing validated empirical support for proposed theoretical advantages of system analysis modeling techniques.
2. One study will utilize the proposed empirical technique to compare simplified forms of "Structured Analysis" and "Object Oriented Analysis". This study will provide empirical evidence of the relative performance of individuals using the object-oriented and structured analysis approach.
3. The empirical instrument developed earlier will be used a second time to compare alternative grammars for developing an Entity Relationship Diagram. This study will provide evidence of the ability of the empirical instrument to compare small differences within a single analysis approach. This study also provides a method for assessing the prediction developed from the ontological approach suggested by Wand & Weber (1990,1995).
4. The proposed empirical instrument can be used to assess two important questions surrounding analysis methods. First, the relative effectiveness and efficiency of text descriptions as compared with graphical descriptions. Second, the effect that knowledge of the diagramming technique and prior knowledge of the domain has on the level of understanding developed by individuals.

1.7 Organization of the Thesis

The remaining chapters are organized in the following manner. Chapter 2 outlines previous empirical comparisons of system analysis methods. The emerging importance of multiple measures and the need to measure both the product and the process of developing an understanding of a business domain are discussed.

Chapter 3 follows this discussion with the empirical and theoretical background for the proposed empirical comparison method. This empirical background is based primarily on the research of Richard Mayer, an educational psychologist, whose work has focused on how individuals develop an understanding through conceptual models. The theoretical background, where applicable, is based on the ontological approach suggested by Wand & Weber (1995).

Chapter 4 uses the empirical and theoretical foundations to develop hypotheses and related predictions that will form the basis for the two studies to be reported in this thesis. These hypotheses are based on the model developed by Mayer along with the ontological approach suggested by Wand & Weber (1993). Four main hypotheses are discussed along with related hypotheses.

Chapter 5 outlines the empirical designs for both studies in this thesis and highlights the experimental procedures used in the studies. Following the discussion of empirical methods, Chapter 6 will discuss the statistical procedures used in the study and the testing of assumptions underlying these procedures. Chapter 7 will then present the results of the empirical studies and discuss the results. Finally, Chapter 8 will present the conclusions that emerged from the analysis of results and briefly mention the limitations of the study and the potential for future research.

Chapter 2: Empirical Background

2.1 Describing Information Systems with Grammars

Information systems researchers have long understood the challenges in describing the domains related to information systems (Gane & Sarson, 1979). The characteristics of these domains – intangible, technical, complex, and overwhelming detailed - make them particularly difficult to describe. The description is made even more difficult when considering the audience for these descriptions. A domain description must be capable of communicating system information to business experts, who generally have little knowledge of system technology, while at the same time communicating technical information to system developers, who may know little about business. What tempts researchers into considering this challenge is the fact that at the heart of every information system is a set of well formalized, ordered, and logical actions which consistently move the system into predictable, abstract, and describable states.

Researchers have responded to this description challenge by developing a large number of system analysis modeling techniques. These analysis techniques, in general, provide a set of modeling constructs for describing an information system along with rules for combining these constructs. The constructs and rules, taken together, describe a **grammar**, like English grammar, that can be used to build system descriptions. In developing comparisons between system analysis techniques, we are in essence comparing the ability of alternative grammars to represent a domain.

Having introduced the idea of alternative grammars, it is natural to consider the questions as to which grammars should be compared and what dimensions, concepts, and measures should be used in the comparison. This study addresses the lack of systematic empirical research by first examining previous empirical research to understand the comparisons of analysis techniques that have already been made.

2.2 Previous Empirical Comparisons

It is unfortunate that only a handful of previous researchers have used empirical information to draw comparisons between modeling techniques[3]. Comparative evaluations have been made on relational and hierarchical database models (Brosey & Schneiderman, 1978); flowcharts versus program design languages (Ramsey, Atwood, & Van Doren, 1983); data flow diagrams and task oriented menus (Nosek and Ahrens, 1986); data flow diagrams and $IDEF_0$ models (Yadav, Bravoco, Chatfield, & Rajkumar, 1988); logical data structure and relational data structures (Jarvenpaa & Machesky, 1989); relational data model and extended entity relationship model (Batra, Hoffer, & Bostrom, 1990); object, process, and data modeling methods (Vessey & Conger, 1994); optional versus mandatory properties in entity relationship diagramming (Bodart and Weber, 1996); and the relationship construct (Siau, 1997).

These empirical comparisons have used a number of measurement instruments to measure a variety of constructs and a variety of grammars. The measurement instruments include timed and non-timed comprehension tests, ease of use questionnaires, ease of learning questionnaires, expert ratings of models developed by participants, along with verbal and behavioral protocols. The constructs measured by these instruments include model accuracy, accuracy of understanding, semantic accuracy, syntactic accuracy, syntactic completeness, ease of use, ease of learning, number of breakdowns (errors or slips in comprehension), and content of knowledge. A summary of these studies including the subjects, measuring instruments, and constructs used in each study is provided in Table 2 below.

The small number of studies and the differences in measurement instruments and constructs used in these studies indicate two things. First, that useful empirical comparisons of systems analysis methods are difficult to create. Much of the difficulty, as noted in Wand & Weber (1993) and indicated earlier by Floyd (1986) has to do with a lack of a theoretical basis for the research. Second, the diversity of instruments and constructs suggests that no widely accepted set of instruments or constructs for model performance have emerged. While no generally accepted constructs for evaluating modeling methods have evolved, the previous studies have shown some convergence on the important issues of *who, what,* and *how* techniques should be measured. These issues are of particular importance in considering further empirical work and are discussed below.

[3] In fact, Vessey and Conger (1994, p.102) note that their literature review revealed only a single study by Yadav, et.al. (1988) comparing methods for requirement specification. A broader search focus for the current study provided only nine studies, further indicating the relative scarcity of empirical comparisons of modeling techniques.

Table 2: Summary of Empirical Analysis Method Comparisons

Authors	Comparison	Subjects	Instruments	Constructs
Brosey & Schneiderman , 1978	Relational and hierarchical models	novices	Comprehension test	comprehension, query accuracy
Ramsey, Atwood, & Van Doren, 1983	Flowcharts versus program design language (PDL)	know-ledgable students	expert rating of product developed in study	design style, level of detail, model constructs, model correctness
Nosek & Ahrens, 1986	data flow diagrams and task oriented menus	novices	Comprehensive test	accuracy of understanding
Yadav, Bravoco, Rajkumar, & Chatfield, 1988	DFD and IDEF0	novices	expert rating of models, ease of use/ ease of learning	semantic correctness syntactic correctness and completeness, ease of use, ease of learning
Jarvenpaa & Machesky, 1989	logical data structure (LDS) and relational data structures (RDS)	novices	timed comprehensive tests. Expert rating of model accuracy, verbal protocols	model correctness, interpretation accuracy, ease of learning, analysis approach
Batra, Hoffer, & Bostrom, 1990	Relational data model (RDM) and extended entity relationship model (EER)	novices	expert rating of representation created in lab, ease of use questionnaire	model correctness, ease of use
Vessey & Conger, 1994	object, process, and data methods	novices	Behavioral protocols, verbal protocols,	number of breakdowns, procedural or declarative knowledge
Bodart & Weber, 1996	Optional and mandatory properties in entity relationship diagrams	novices	Free recall and comprehension	Number of items remembered
Siau, 1997	Explicitness of relationship and noun vs. Verb description	novices	Verbal protocols	"chunks" correctly interpreted

Who Should be Measured?

The most common participants used for the comparisons were novices. Ramsey, Atwood, & Van Doren (1983) was the only study that tested individuals with a high level of understanding of the modeling method. Vessey & Conger (1994, p. 104) argued that novices are useful subjects when comparing analysis techniques for two reasons:

> *"First, ... it is easier to teach them (novices) to apply a specific methodology than it is to teach new methods to people who may already be experts in developing systems... Second, examining expert problem solving can be quite difficult, since experts automate their processes to the point at which they are no longer able to articulate what they are doing. Novice problem solving, on the other hand, is much more amenable to investigation "*

Arguing for external validity, Batra, Hoffer & Bostrom (1990) suggest that novices are representative of many end-users of information systems who are involved in the analysis process. The results obtained from novices are, therefore, more likely to indicate the amount of information communicated between individuals in many business settings. Another argument for using novices is that system analysis experts are likely to already have established a method for describing systems. The experts are more likely to be biased towards the system analysis technique that they use in practice and therefore do not make attractive participants. While not stated, another practical argument for the use of novices in empirical studies is their relative abundance in comparison with the small number of available experts.

While novices were used often, many of the researchers indicated that extensions of their experiment to experts are important. Brosey & Schneiderman (1978, pp. 634) indicated that "*Testing of professional or clerical staff ... would be useful for comparison,*" while Vessey & Conger (1994, pp. 112) stated " *similar studies should be undertaken with experienced system analysts to assess their performance.*" Batra, Hoffer, & Bostrom (1990, pp. 137) also suggested that "*Another extension of this research would be to compare expert and nonexpert designers, so that one could identify the nature of expertise in this context.*" These statements suggest that while comparisons based on novices provide useful information, testing experts provides a further degree of external validity while delivering data on the differences between expert and novice cognitive models. Of course, as mentioned earlier, the bias of experts towards familiar methods needs to understood and accounted for.

What Should be Measured?

Table 2 provides a variety of constructs that researchers have identified as important in measuring the relative performance of alternative modeling methods. None of the authors agree completely on the constructs used in developing empirical comparisons. Recent studies have converged, however, on the importance of assessing both the ***process*** that an observer uses to develop an understanding, as well as the ***product*** of understanding. The "process" represents the cognitive activity necessary to create an understanding of the domain being represented. This includes searching, integrating, and storing the information about the domain. The "product" of understanding is the cognitive model that an individual has developed as a result of viewing the diagram or description. Since the product is cognitive, the product cannot be adequately observed directly. For this reason, participants are asked a set of performance tests, to provide a proxy for the "product" that each participant has developed cognitively.

Yadav et. al. (1988) were first to explicitly identify and measure the separate dimensions of modeling process and product, a distinction that is clearly made in a related paper from Vessey, Jarvenpaa, & Tractinsky (1992). The early empirical comparisons in Table 2 (Brosey & Schneiderman (1978), Nosek & Ahrens (1980), Ramsey, Atwood, & Van Doren (1983)) did not identify the process dimension, focusing exclusively on the product of modeling techniques. These studies relied on comprehension tests or expert ratings of developed models to provide comparative measures. They collected little or no observations on the process of developing an understanding of the domain being represented.

Later studies increased the attention focused on the process of understanding. Jarvenpaa & Machesky (1989) and Batra, Hoffer, & Bostrom (1990) both measured the dimensions of process and product. Besides comprehension questions to measure the product, both studies measured the ease with which understanding was developed using either an ease of use questionnaire (Batra et. al., 1990) or a timed comprehension test and analysis of verbal protocols (Jarvenpaa & Machesky, 1989). In both of these studies, the primary focus of the empirical comparison continued to rest, however, with measurement of the modeling product.

In Vessey & Conger (1994) and Siau (1997) the focus shifted from observing the *product* produced while modeling, to observing the *process* of developing understanding from a model. To capture information regarding the process of understanding, verbal protocols of participants were observed. The emphasis placed on the process of understanding by Vessey & Conger (1994) underlines the importance of process in making empirical comparisons of modeling methods. Siau (1997) managed to capture both process and product information in the protocol method applied. The large number of protocols collected by Siau (1997) enabled a comparison between participants groups that provided an added objective dimension to the more "subjective" method of protocol analysis. Previous empirical studies indicate, therefore, that a full comparison of modeling methods requires the observation of both the final product of understanding (an individual's cognitive model) as well as the process required in developing that understanding.

Another important distinction arising from a comparison of the studies in Table 2 is the tasks that the studies are focused on. There are two basic tasks in any modeling exercise: representation (writing the model) and interpretation (reading the model). The studies listed in Table 2 were split on the task that was evaluated. Brosey & Schneiderman (1978), Nosek & Ahrens (1986), Bodart & Weber (1996), and Siau (1997) chose to observe the process of interpretation. To do this, these researchers created a representation and the asked participants to view the interpretation and answer questions about it. On the other hand Ramsey, Atwood, & Van Doren, (1983), Yadav et.al. (1988), Batra, Hoffer, & Bostrom (1990), and Vessey & Conger (1994) all chose to observe representation. In all cases, except Vessey & Conger, the observations were made on the end product of representation, which was the final diagram. In general, experts were used to rate the diagram's accuracy. Vessey & Conger focused on the representation process by collecting verbal and behavioral protocols while the participants drew the diagrams. Jarvenpaa & Machesky (1989) was the only study to capture data on both the representation and interpretation tasks.

One final notable distinction between the studies is the level of analysis. A majority of the studies focused on analysis methods that were closer to implementation and design than to the analysis and conceptual models. For example, Brosey & Schneiderman (1978), Ramsey, et. al. (1983) Nosek & Ahrens (1988), Jarvenpaa & Machesky (1989), and Batra et.al. (1990) used diagrams that presented a relatively "low" level analysis of a domain. These types of models were created primarily to communicate domain details with developers. On the other hand Yadav et. al. (1988), Vessey & Conger (1994), Bodart & Weber (1996) and Siau (1997) used "higher" level models that were farther removed from the implementation models. While both levels of modeling are important, the communication to the user of a system is more likely to occur with "high" level models. It is important, therefore, to consider the potential audience that the model will be addressing when designing the experiment.

How Should a Technique's Performance be Measured?

The number and types of instruments for measuring model performance varied widely across the empirical studies. This is both unfortunate and understandable. The different measurement instruments reflect the different constructs used by each of the authors. As each measurement instrument has inherent strengths and weaknesses, instruments are more likely to be used on constructs that reflect the instrument's strength. For example, the nature of verbal protocols suggests that they are more likely to be used to gather information on processes, and less likely to be used to gather information about the product. Of course, a combination of methods is possible where a researcher can collect protocols while asking a participant to complete a performance task that measures product as well.

The changes in constructs between authors may also reflect the difficulty with developing appropriate measurement instruments. Constructs associated with modeling technique performance are not amenable to quick, easy to use, objective measurement instruments. The difficulty arises from the nature of cognitive processes in individuals with different experiences, memory processes, cognitive styles, and levels of motivation. The challenge in developing an empirical comparison is to provide a set of observations that affords a reasonable level of both internal and external validity.

While the previous authors may not agree completely on a set of measurement instruments that provide convincing information on modeling technique performance, some convergence can be observed. For example, Yadav et. al. (1988), Jarvenpaa & Machesky (1989), Batra, Hoffer, & Bostrom (1990), and Vessey & Conger (1994) all utilize more than one measurement instrument in developing their comparisons. This suggests that measuring technique performance may require more than a single instrument. Comparisons that utilize a single measurement instrument run the risk of missing important comparative information. Different measurement techniques can also provide the opportunity for construct validation provided by the "triangulation" of results. A further convergence between the studies can also be observed on the use of verbal protocols for measuring the process of developing understanding. This can be seen in the measurement instruments described in Jarvenpaa & Machesky (1989), Vessey & Conger (1994), and Siau (1997).

2.3 Summarizing Insights from Previous Research

The eight studies summarized in Table 2 above provide useful insights into who, how, and what should be measured when considering an empirical evaluation of analysis techniques. Novices were the overwhelming choice for subjects, however, authors have urged that future research be extended to experts. When considering what should be measured, the emerging consensus suggested that both the product and processes associated with modeling technique performance should be measured. Further, these processes and products should be compared through a variety of empirical measures since no single measure can be used to capture the understanding of a person viewing a system model.

In the next chapter, the theoretical and empirical foundations will be laid for the development of an empirical instrument for making comparative evaluations of alternative system analysis methods.

Chapter 3: Empirical and Theoretical Foundations

The discussion of background for the study is divided into two sections: empirical and theoretical foundations. The first section of the chapter will consider the foundations underlying the empirical techniques used in the studies. The second section then outlines the theoretical foundations on which some of the hypotheses regarding the empirical comparisons will be based. These sections will be used to develop hypotheses and working hypotheses in Chapter 4.

3.1 Empirical Foundations

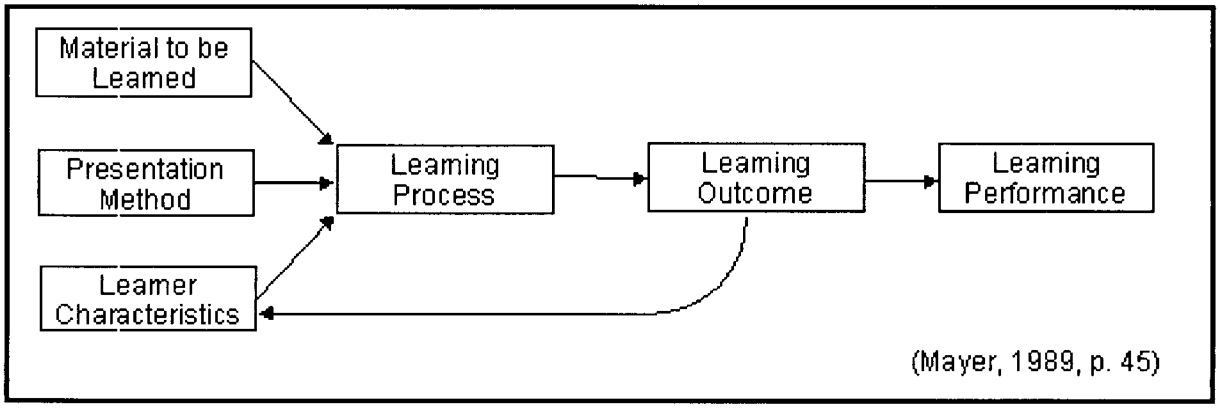
The empirical foundations begin with a definition of the area of research interest. In this study, analysis methods will be compared on their relative ability to capture and convey **domain** information, and not on their ability to convey information about the **information system** to be implemented. This suggests an implicit bias towards the early stages of systems analysis, often referred to the area of "**Conceptual Modeling**". A useful definition for this modeling activity is provided by Mylopolous (1992, p. 2), who suggests:

> "*Conceptual modeling* is the activity of *formally* describing some aspects of the physical and social world around us for the *purposes of understanding and communication."*

Viewed from this perspective, the early phases of system analysis can be characterized as a **learning process** where materials are **communicated** to individuals in an effort to improve their **understanding** of the domain under consideration.

Having characterized conceptual modeling as a learning process, the discussion of how to assess the learning process can be organized by making use of the model of learning suggested by educational psychologist, Richard Mayer (1989). The learning model contains six components as shown in Figure 1. Three of the components are antecedents for the learning process (learning material, presentation method, and learner characteristics). The three remaining components identify the learning process, learning outcomes, and learning performance. Each of these components, in the context of this study, is described in more detail below.

Figure 1: Components of Learning Process



Material to be Learned

The learning material (or material to be learned) in this study are descriptions of organizational processes. Example processes in this study include organizing an academic conference, organizing an entertainment event, organizing a bus company, and organizing a machine repair facility. Examples of the text descriptions used in this study are provided in Appendix A.

Presentation Method

The method used to present the domain description to the person viewing the model is referred to as the **presentation method**. Four presentation methods will be used in the two studies: text description (TXT), object-oriented diagrams (OOD), data flow diagrams (DFD), and Entity Relationship Diagrams (ERD).

Each presentation method relies on a **grammar** from which a model of a business domain is developed. A grammar is comprised of a set of descriptive **symbols** and a set of **rules for combining** these symbols. For example, the English language has a set of symbols (letters of the alphabet and punctuation marks) that can be combined using a set of rules to present information about a domain. Together these symbols and rules make up the English language grammar. Data flow diagrams, in a similar way, also have a set of symbols (process, data store, data flow, and external entity) that can be combined using a set of rules to produce a graphic representation of a process. This set of symbols and rules for acceptable combinations of symbols make up the grammar underlying the data flow diagram.

The comparison of two presentation methods, such as text and data flow diagrams, is essentially the comparison of the effectiveness of different grammars in communicating relevant domain information (Wand & Weber, 1993). This study will control the type of grammar (presentation method) used to present domain information, and then attempt to measure the amount of understanding that individuals develop about a particular domain while viewing the model.

Learner Characteristics

Two characteristics of the learner (person viewing the model) are immediately important in this study: prior experience with the domain, and prior experience with the presentation method. The prior **experience with the domain** may influence the level of understanding attained when viewing a model of the domain. All things equal, the more domain experience an individual possesses, the higher the expected level of domain understanding. The prior **experience with the presentation method** can also affect the level of understanding when viewing a representation. Again, all things equal, the more competence an individual has with the presentation grammar, the higher the expected level of domain understanding after viewing the representation.

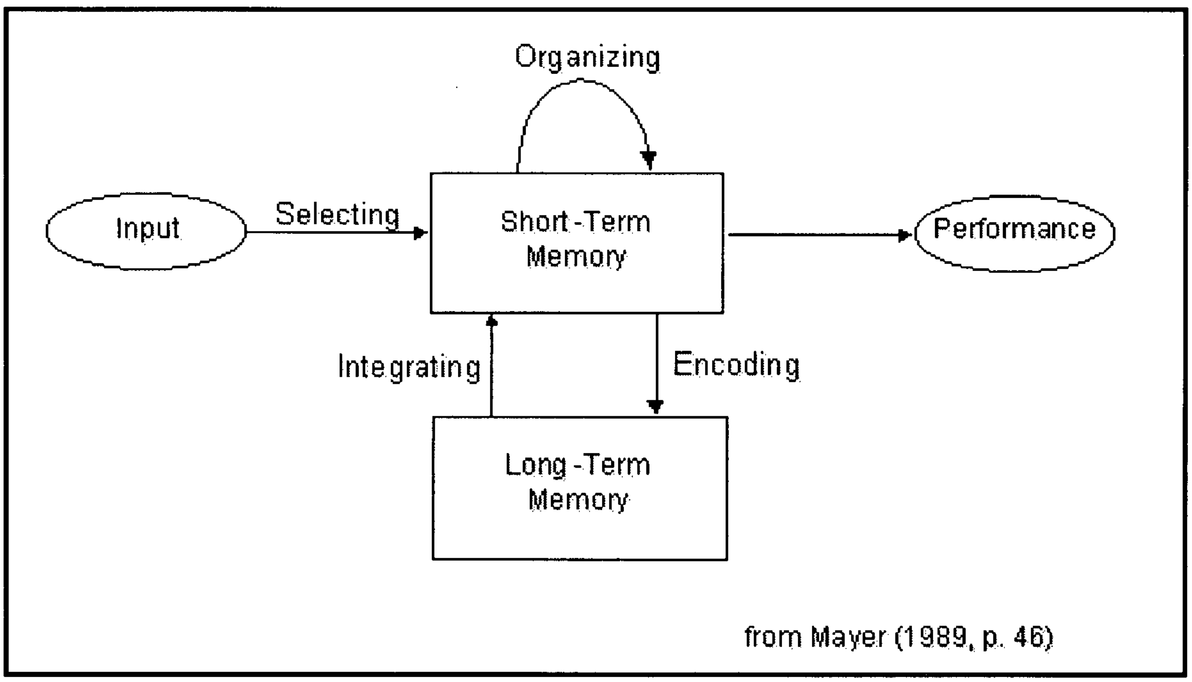
Given the influence of these factors, it will be important to attempt to measure both the level of domain experience and the experience with the presentation method, before observing learner performance. In an effort to reduce the influence of domain experience, cases used in this study have been chosen for their prima facie lack of familiarity for the population under study. Assessing the prior knowledge regarding a domain is difficult as the assessment can change the actual level of awareness before the test is begun. For this reason no test of the familiarity of the domains used in the study were conducted. A simple measure of domain knowledge was collected in the pretest as described in Chapter 5. In dealing with presentation method experience, this study incorporates differences in modeling competence directly into the study design. The study will seek out participants who have attained some competence with the modeling technique as well as individuals who have no experience with the modeling technique. By choosing an appropriately large sample size and randomizing individuals across groups, the confounding effects of domain experience and presentation method experience can be reduced.

Participants without prior modeling experience will provide a potentially valuable control group from which comparisons can be drawn with participants who have modeling experience. These comparisons may be especially important in providing information about how system "users" - who are often portrayed as individuals with little or no modeling experience - develop understanding from diagrams created from grammars of which they have little or no experience.

The Learning Process

The learning process refers to the way in which participants will interpret and enter domain information into short and long-term memory. Mayer's (1989) model of human information processing, which has been adapted from Simon and Newell's (1972) Human Information Processing model will be adopted to explain these processes. This model is shown below in Figure 3.

Figure 2: An Information Processing Model



The boxes in Figure 3 refer to memory (short and long term), arrows represent cognitive processes, and ellipses indicate input or output of these processes. Mayer suggests that well designed representations (conceptual models) affect the learning process by helping individuals **select** appropriate information, and by helping to **organize** the information in short term memory. The selected information in short-term memory can then be **integrated** with prior knowledge from long term memory. This results in a learned outcome that can then be

encoded into long term memory. If the conceptual model design is effective, the cognitive model will be improved and the new information encoded into long term memory provides the model viewer with the potential for improved **learning performance**.

The direct observation of the learning process is both important and difficult. The observation of the learning process is important because analysis methods that are easy to learn, and easy to learn from, are preferred to those that are more difficult to learn (all other things equal). For this reason, some observations regarding the ease with which the method can be applied should be taken. Unfortunately, observation is difficult since the learning process is not directly observable.

One method for opening the learning process to observation is verbal protocol analysis (Ericsson & Simon, 1993). In this observation method, participants are encouraged to "think aloud" as they move through the learning process. The researcher is thus provided with a running monologue of "thoughts" that occur through the learning process. While verbal protocols provide a rich data source, they tend to be difficult to analyze and are normally attempted with a small number of participants. A further discussion of verbal protocols is left to chapter 5.

A less direct method for observing the learning process is provided by asking participants to estimate their perceived "ease-of-use" associated with a modeling technique. An instrument for collecting this perceived measure has been developed by Moore & Benbasat (1991). While originally designed to measure the ease of use of information technology, this study will suggest that the instrument can be extended to analysis methods. The ease of use scale provides a measure for the learning process that a participant goes though to develop understanding about a domain. The short form (four questions) scale was adapted for this research. The questions related to ease of use are provided in Appendix F.

Learning Outcome: Domain Understanding

When viewing system analysis models, the learning outcome is the amount of domain knowledge that the person acquires (encodes) as a result of viewing the model. Mayer (1989, p. 47) suggests that when compared to participants viewing text descriptions, persons viewing conceptual models,

> "...*may be more likely to build mental models of the systems they are studying and to use these models to generate creative solutions to transfer problems. In short, these students may be better able to engage in systematic thinking.*"

While the learning outcome is not directly observable, Mayer suggests that a positive learning outcome occurs when conceptual models are used to enhance the selection, organization, and integration processes required for developing understanding. He also suggests that positive learning outcomes are revealed through superior learner performance.

Learning Performance: Comprehension, Recall, Problem Solving

Learning performance refers to how a participant, as a result of learning, performs in tasks that require domain knowledge. To measure learning performance this study will make use of some of the measures described in Mayer (1989). A short description of these measures is provided below.

Mayer began his experiments by identifying two treatment groups: one group was provided with a text description accompanied by a diagram ("model" group) and one provided only with text description ("control" group). He then suggested that exposure to a diagram would improve the quality of the cognitive model developed by subjects (students). After the subjects viewed the material, Mayer asked them to complete three tests: comprehension, verbatim recall, and problem solving.

The comprehension tests included questions regarding the attributes of things or the relationship between things in the explanative material. For example, in Mayer & Gallini (1990) participants were provided with information on the braking system of a car. Comprehension questions included questions such *as "What are the components of a braking system*" or "*What is the function of a brake pad*".

After the comprehension tests, participants were given a set of questions that went beyond the original explanative description provided. These types of questions Mayer referred to as ***problem solving*** tasks. The idea behind the problem solving task was that individuals that are able to form "better" cognitive models of the material presented to them will provide a larger number of correct solutions to the question than individuals with less well-formed cognitive models. The questions are designed so that there is more than one correct answer. An example of a problem solving task in the Mayer & Gallini (1990) study included questions such *as "What could be done to make brakes more reliable*?" or " *What could be done to reduce the distance needed to stop*."

Finally, after the problem solving task, the participants were given a verbatim recall test. The verbatim recall test provided each participant with a collection of paired statements. One statement was taken verbatim from the text description. The other statement was altered slightly from the verbatim statement. The participant was then asked to select the statement that was exactly the same as the statement that was given in the original description.

Mayer made three predictions regarding these three tests. First, "model" and "control" subjects will generally have similar scores in comprehension tests. The second prediction was that "model" subjects score lower in verbatim recall than control students. The lower scores reflect the fact that the model students reorganize and integrate information due to the impact of the model, and are therefore less likely to retain information in verbatim form. Finally, Mayer (1989) predicted that "model" subjects would generate a larger number of creative solutions to problem solving tasks than control students. This prediction suggests "model" students possess more sophisticated cognitive models from which to generate creative solutions than control students. This third prediction was the most important for Mayer; higher scores in problem solving would indicate a more refined cognitive model. Mayer (1989) provides a good summary of the success of his measurement techniques and the related predictions. In general, all three predictions were substantiated in a large number of samples and situations, so the measure provided a high degree of internal validity.

3.2 Adapting Mayer's Instrument for System Analysis Methods

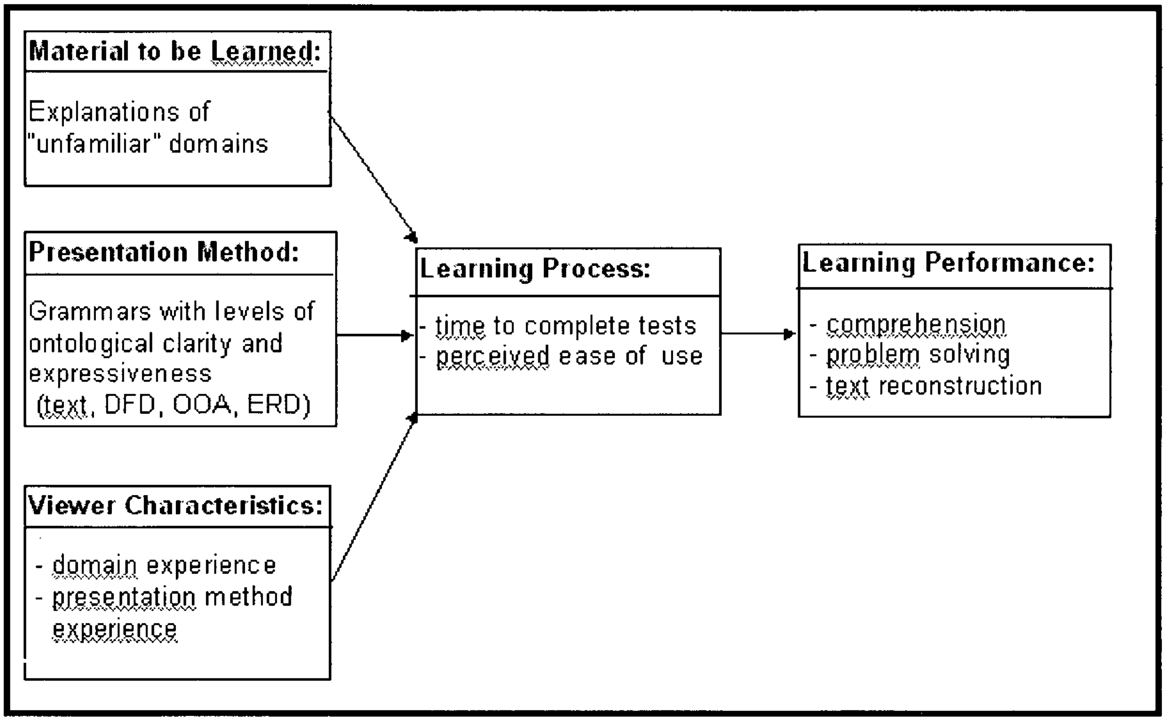
In adapting Mayer's methods for the comparative measurement of analysis methods, the comprehension and problem solving tests remain largely unchanged. The verbatim recall task, however, is not directly transferable as the presentation methods proposed for this study make use of both text and graphic objects. The verbatim recall task has been replaced with a text reconstruction task. In this task, the participant is provided with the text originally used to create the representations, but from which important words have been replaced with blank spaces. The text descriptions were adapted from cases in Bodart & Weber (1996), Batra, Hoffer, and Bostrom (1990), and the IFIP 8.1 Working group case. The researcher created one of the cases. In this test the important words were chosen by the researcher, but the words are often eliminated either randomly or in a pattern (for example, eliminate every fifth word). The participant is encouraged to fill in the blank spaces with the word they believe best “fits”. In this way, the participant reconstructs the original text. This type of "fill-in-the-blank" test is referred to as a "Cloze" test, as described by Taylor (1953).

The Cloze test was chosen for several reasons. First, the test is a different measure of understanding when compared to short answer and multiple choice comprehension tests. McKenna & Robinson (1980) indicate that the Cloze test measures comprehension, and Cloze scores are positively correlated with comprehension scores. The Cloze test is different from comprehension, however, because the score depends less on short-term memory and more on semantic structure and conceptual understanding (Bormuth, 1967). Second, the more random nature of the deletion of text, makes the Cloze test less likely to be biased by the researcher developing the test than comprehension questions. Third, the Cloze test provides an effective approach to examining individuals concurrently interacting with visual information such as diagrams (Taylor, 1953). Finally, the Cloze test also provides a source of locative information, pointing towards sections in the text that are particularly difficult for the participants. While the Cloze does not replicate the function that a verbatim recall test provides, the Cloze test does provide another useful measure of learner performance.

3.3 Overview of Empirical Components

The discussion of empirical foundations is summarized in Figure 4 below. The figure expands the model provided earlier in Figure 1 to include a listing of the constructs and measurement methods to be utilized in this study. A brief overview of Figure 4 shows three dependent measures of learning performance: comprehension, problem solving, and text reconstruction. Two indices of the learning process – time to complete task and perceived ease of use – are also collected. Each of the dependent factors are influenced by three independent factors: the grammar (presentation method) used to present the material, the level of previous domain knowledge held by each participant, and the level of knowledge of the presentation method (analysis technique grammar).

Figure 3: Overview of Empirical Components



Two studies are proposed based on this description of the learning process. The first manipulates presentation method and learner characteristics to compare the effects on learning performance on three treatment groups: text description, a combination of a Data Flow Diagram and Entity Relationship Diagram, and Object Oriented Diagram. The second study again manipulates the presentation method to revisit the study reported by Bodart & Weber (1996) based on different grammars for the Entity Relationship Diagram. These studies are detailed in the following chapters.

3.4 Theoretical Foundations

The lack of a theoretical basis for the comparison of system analysis methods is well documented. Researchers such as Colter (1984), Floyd (1986), Yadav et. al. (1988), and Wand & Weber (1993), have noted the lack of theory and have suggested frameworks or sets of concepts upon which the comparisons of information system analysis (ISAD) grammars could be based. These suggested sets of concepts have not been explicitly recognized by later empirical research (Green, 1996). The lack of accepted measures for empirical work reveals the lack of a theoretical base for the comparisons (Wand & Weber, 1995). This problem is clearly identified by Floyd (1986, p.31) who noted:

Unfortunately, our understanding of the nature of system development as a whole is haphazard... and tends to be based on opinions and individual experiences rather that on systematic empirical research.... the lack of a suitable theory about system development as a whole explains many of the shortcomings in the existing methods."

It is through the development of a theoretical basis for analysis methods that further empirical advances can be made.

Ontology as a Basis for Comparison

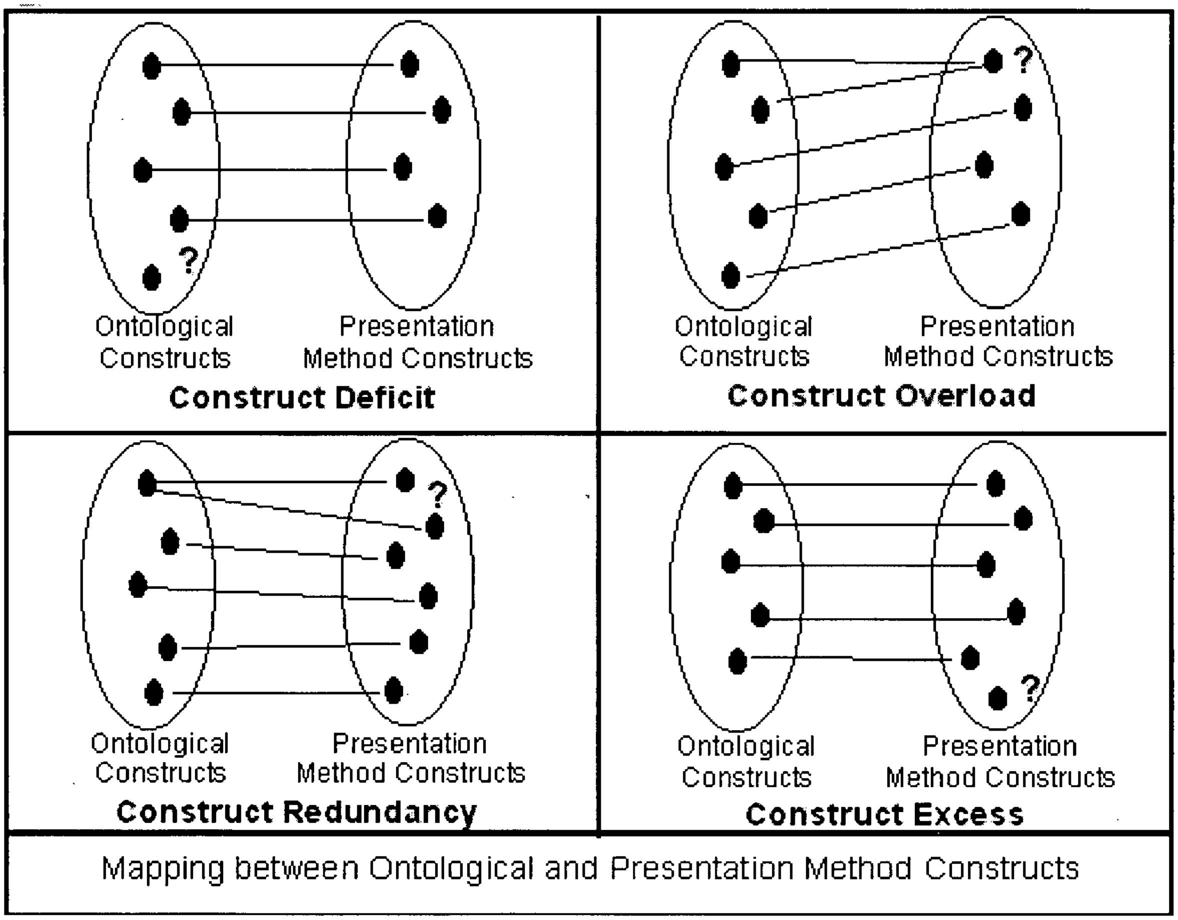
Wand & Weber (1990, 1993, and 1995) have addressed the lack of theory by suggesting the use of ontology as a theoretical basis for analysis methods. Ontology is the branch of philosophy concerned with how persons describe the world around them. Wand & Weber (1990, 1993, 1995) propose a set of ontological constructs, based on the work by Mario Bunge (1977,1979), that provides a standard ontology upon which alternative grammars can be compared. Bunge's ontology was chosen for several reasons as noted by Wand & Weber (1995):

"We have chosen to work with Bunge's ontology because it deals directly with concepts relevant to information systems and computer science domains (e.g. systems, subsystems, ad couplings). Moreover, Bunge's ontology is better developed and better formalized than any other we have encountered."

The ontological constructs described in the BWW (Bunge, Wand and Weber) model provide what they suggest is a complete set of constructs for describing information system domains. These constructs are described in Appendix I.

In addition to their proposed ontological constructs, Wand & Weber (1990, 1993) have also identified four situations that can occur when modeling grammars are compared with the BWW ontological constructs. These concepts are construct deficit, construct overload, construct redundancy, and construct excess. At the heart of these concepts is the mapping of the proposed ontological constructs onto the constructs used in the grammar (presentation method). For example, data flow diagrams describe the world using four basic constructs - process, data flow, external object, and data store - along with a small set of rules for combining these constructs. To assess the DFD grammar in terms of the BWW ontology, one must look at how the set of DFD constructs map onto the set of ontological constructs proposed in the BWW ontology. Four general types of mappings are shown below in Figure 4. These situations are important in developing hypotheses regarding the comparative performance of analysis methods. Each of these situations is discussed briefly below.

Figure 4: Mapping between Ontological and Presentation Method Constructs



* Diagram adapted from Wand & Weber (1993, p. 223)

Construct Deficit

A presentation method grammar has ***construct deficit*** if there is at least one ontological construct that does not map into at least one presentation method grammatical construct. This type of mapping suggests that the presentation method is ***ontologically incomplete*** because the presentation method cannot represent at least one necessary aspect of reality (ontological construct).

Construct Overload

A presentation method grammar has ***construct overload*** when one presentation method construct maps into more than one ontological construct. This means that the presentation method is ambiguous in that it uses one construct to map into two separate ontological constructs.

Construct Redundancy

A presentation method grammar is in ***construct redundancy*** when two presentation constructs map onto the same ontological construct. This means that the presentation method is again ambiguous. As with construct overload, this ambiguity will affect the ***ontological clarity*** associated with the presentation method grammar.

Construct Excess

A presentation method has ***construct excess*** when the presentation method grammar has a construct that does not map into any ontological construct. This means that the presentation method is again ambiguous because the presentation method grammar has too many constructs. As with construct overload and construct redundancy, the excess construct will affect the ***ontological clarity*** associated with the presentation method grammar.

The process that Wand and Weber (1990, 1993, 1995) have outlined suggests that system analysis method constructs (i.e. presentation methods constructs) should first be analyzed using the ontological constructs as a basis for the analysis. Predictions on the performance of the grammars can then be based on whether one or more of the mapping situations described above arises from the analysis. For example, given that a first presentation method has construct deficit and a second method does not, Wand and Weber would predict that the second method would outperform the first method. What is important in this comparison is that the hypothesis is based on a strong theoretical foundation and can be tested empirically.

3.5 Linking Ontological Constructs to Cognitive Constructs

In analyzing Wand and Weber's proposed method for comparing system analysis methods, it is important to understand that Bunge's ontology is not the only ontology available for the purpose of describing information systems. Wand & Weber (1990, 1993, 1995) argue, as noted earlier, that Bunge's ontology is useful because it is rigorously defined using set theoretical language that could be readily adapted to information systems. Wand and Weber recognize that alternative ontologies exist. No theoretical proof can be offered to show that Bunge's ontology is a better platform than other ontologies designed for information systems.

The ontology must be assessed using empirical data to see if the constructs as defined by Wand and Weber are useful.

The argument that grammars with – for example - construct deficit as defined in the BWW ontology, are less effective than grammars without construct deficit makes an assumption regarding individual cognition. The assumption is that the BWW ontological constructs closely mirror the cognitive constructs that individuals use to think about information systems. If individuals use constructs similar to the BWW ontological constructs to think about systems, we could then predict that grammars with mapping anomalies, such as construct deficit or construct redundancy, will lead to a less well developed understanding than grammars without the mapping anomalies. This implies that when we assess predictions based on the analysis using the BWW model, we are actually assessing to see if the ontological constructs proposed in the BWW ontology are close to the cognitive constructs people use to think about systems. Until we have developed a theoretical basis for linking ontological constructs to cognitive structures, the degree of relationship between cognitive constructs and ontological constructs will remain an empirical issue.

Some empirical evaluation of the ontological constructs has been undertaken. Weber & Zhang (1991) analyzed Nijssen's Information Analysis Method (NIAM) and made predictions developed on the basis of that analysis. While the number of subjects was small, and the empirical technique was not well defined, some indication for the usefulness of the BWW ontological constructs in predicting problem in an analysis method did exist. In a later paper, Sinha & Vessey (1995) used the BWW constructs to evaluate relational and object oriented schema diagrams. The results in this test were inconclusive. Bodart & Weber (1996) also used the BWW ontological constructs to suggest differences in the understanding developed from alternative entity relationship diagramming grammars. This test will be described in more detail in the chapters that follow. Again, in this case, the results were not conclusive.

3.6 Focus on Interpretation

Two activities are associated with the development of any domain description. Descriptions must be created and descriptions must be read. As noted earlier, r**epresentation** is the activity in system analysis within which the description of a domain of interest is created. **Interpretation** is the activity in analysis within which the description of a domain is read and, hopefully, understood. A complete evaluation of a system analysis method should consider both the representation of, and the interpretation from, a domain description. This study will focus exclusively on the interpretation of domain descriptions. Further empirical work will have to be done to assess the effectiveness and efficiency with which representations are created with different analysis methods.

The focus on interpretation in this study is justified for two reasons. The objective of this study is to measure the understanding that can be **communicated** through alternative method grammars. Interpretation is the primary activity associated with developing this understanding and is necessarily the most important process for communication. For example, if a diagram is created but it cannot be read, then no information is communicated. So the ability to interpret from a description created using an analysis technique is of primary importance.

From an empirical perspective it also is necessary to make a choice between evaluating either representation or interpretation. This choice is necessary because the combination of these two activities in a single experiment presents potential confounds in experimental design. For example, the act of creating a representation may affect a person's approach to interpreting other representations. Focusing this study on interpretation implicitly indicates that it is not necessarily the ideas as they are ***represented*** on paper or computer screen that are important in the communication of system information, but rather the cognitive model that is created in the mind of the person ***interpreting*** the model. The ability to represent a situation with a grammar does not automatically imply that a person viewing the model will understand the situation as presented.

3.7 Summary

This chapter has established the empirical and theoretical foundations upon which the studies in this thesis are based. Mayer (1989) provides the primary foundation for the empirical methods used in the studies. The adaptation of Mayer's empirical methods to the comparison of information systems analysis grammars also suggested the use of the “fill-in-the-blank”, or Cloze test, established by Taylor (1953). These empirical measures will be used to test hypotheses described in the next chapter. The ontological mapping from alternative analysis grammars onto the BWW ontological constructs proposed by Wand and Weber (1989,1990,1993) is used in the second study. Some empirical studies have already emerged from the BWW ontological approach. The second study focused on entity relationship diagrams will extend previous empirical research based on the BWW ontology with a set of new empirical tools adapted from the work of Mayer (1989).

Chapter 4: Research Hypotheses

This chapter will develop the hypotheses related to two studies. The first study focuses on a comparison between grammars that are used in describing business domains for use in system analysis. These grammars are text descriptions, "structured" analysis (using data flow and entity relationship diagrams) and object-oriented analysis. This first study is labeled an ***intergrammar*** comparison because more than one major grammatical approach to the description of a domain will be compared. The second study will focus within a single grammar. Two alternative entity relationship modeling grammars – one using mandatory properties with subclasses, the other using optional properties – will be compared along with text descriptions. The second study is labeled an ***intragrammar*** comparison as it primarily compares alternatives within the same grammatical approach. The two studies will serve as a test for the sensitivity of the empirical method proposed in the previous chapters.

The development of research hypotheses in both studies begins with description of the task being considered as the basis for both studies. After defining this task, the alternative grammars for each study are considered. After a description of these grammars is provided, explanations for expected differences across empirical measures are developed. Hypotheses are then formed and then several hypotheses are then developed. Methods to test these hypotheses are developed in Chapter 5 and the results of these studies are presented in Chapter 6 and 7.

The Task Underlying the Empirical Comparisons

One of the objectives of this thesis is to compare alternative system analysis method grammars. An important consideration, given this objective, is to understand what analysis grammars are designed to do. Analysis grammars have four primary roles in information system development as noted by Wand et.al. (1995, p. 285). These roles are:

1. Provide a way for developers and users to communicate
2. Increase an developers own understanding
3. Serve as a basis for design.
4. Serve as a document of the original state of a system for the purpose of maintenance.

This thesis will focus attention on the first role in the list above. In other words, we have directed attention to how a system analysis grammar can be used to ***communicate a conceptual model of a domain to a user viewing a description***. There are several reasons for focusing on the function of communicating system information to users. Practitioner surveys, noted earlier in Chapter 1, have clearly indicated that the poor communication between users and system developers is perceived as an important factor in system development success and failure. This communication includes the validation of system specifications and clear specification of system requirements. Since these steps in the development cycle are crucial for project success, the communication of a domain to individuals with little modeling experience is an important factor in system development success. System analysis grammars play an important part, therefore, in the system development process.

Having identified the function of the grammars being observed, another concern regarding the empirical comparisons is to ensure that “interesting” comparisons of alternative design grammars are made. What is meant by an “interesting” comparison? If two grammars are designed for different modeling tasks - for example data flow diagramming (DFD) for modeling the flow of data through a system, and entity relationship diagram (ERD) for describing the static relations between data - an empirical comparison would be of little interest. Empirical methods are not needed to suggest that the DFD would describe processes better than ERD, and vice versa for static data relations. Clearly, the methods chosen for analysis should be designed for the same modeling task. An interesting comparison should also compare methods that are of “interest” to the audience for the research. For this reason, grammars that are widely used either in practice or in research would make for an interesting comparison. The following sections will outline the choices made in developing empirical comparisons in this thesis.

4.1 Study 1: Intergrammar Comparison

The task chosen for this study, as noted above, is to communicate a conceptual model of a domain to a person viewing a description. The alternative grammars chosen to create these descriptions are text descriptions, “structured” analysis using DFD combined with ERD[4], and object-oriented analysis (OOA). The three alternative analysis methods represent three dramatically different approaches to system analysis. The first analysis grammar – text description – represents the earliest attempts at describing systems. The ambiguity associated with text descriptions spurred the development of simple graphical grammars such as system flow charts.

[4] The reason for combining the Dataflow diagram and Entity relationship diagram will be provided in the following section.

Hardware, database systems, and programming languages developed early in the application development area generally separated data organization from data processing. With this split, it was natural that system analysts developed separate tools for analyzing data organization and data processing. The most popular analysis method for data organization was (and still is) the Entity Relationship Diagram (ERD). For data processing the most popular tool was the Data Flow Diagram. "Structured Analysis" grew from a core of data flow diagrams, and then expanded to include the ERD. Proponents of Structured Analysis suggested that these two fundamental approaches (ERD and DFD) should be integrated to provide a complete picture of an application (data organization and processing).

Proponents of object-oriented analysis methods (OOA) have recently challenged the efficacy of "Structured Analysis". Proponents of object-oriented methods have suggested that the object model is easier to understand and more "natural" to use. Very little empirical research has been done to assess these statements. This study begins to address this problem. The differences between these approaches and the need to combine certain methods (for example the DFD and ERD) are explained further in the section that follows. While many different extensions to the basic ERD, DFD, and OOA model have been suggested by researchers, I have chosen to keep the definition of the grammars in this study as simple as possible. The simplicity enables the examination of the core elements of these three grammars and will provide a base for further research into suggested extensions for each grammar. The reader is introduced to each of the methods in the section below.

4.1.1 An Introduction to the Analysis Methods Used in the Intergrammar Study

Text Descriptions

The original analysis techniques began as text descriptions of information system domains. Text descriptions are obviously the most expressive grammatical form. While expressive, text descriptions have some limitations. Gane & Sarson (1976, p. 4) detailed the limitations of text descriptions of information systems:

> *"Since we have had no way of showing a tangible model, we have had to build the next best thing, which is to use English narrative to describe the proposed system. Can you imagine spending five years' salary on a custom built house on the basis of an exhaustive narrative description of how the house will be built? ... If you use English to describe a complex system... the result takes up so much space that it's hard for the reader to grasp how the parts fit together. Worse than that... English has some built in problems that make it very difficult to use where precision is needed."*

The shortcomings of text descriptions spawned the development of graphical techniques for analysis such as flow charts, entity relationship diagrams (ERD), and logical data flow diagrams (DFD). Two of these graphical techniques, ERD's and DFD's, developed into widely accepted norms for representing process and data structures, respectively, in information system domains. Green anecdotally noted (1996, p. 193) the acceptance of these techniques in his study of Computer Aided Software Engineering (CASE) users:

> *"... logical data flow diagrams (LDFD) are most commonly used (96.7percent), followed by entity relationship diagrams (ERD) (61 percent), data model diagrams (DMD) (53.7 percent), and structure charts (STC) (38.2 percent). "*

This study makes use of basic forms of the ERD and logical DFD as provided, respectively, in Chen (1976)[5] and Gane & Sarson (1979). It is recognized that extensions to these basic grammars are possible and popular. This study, however, is interested in identifying differences between the core concepts of the three grammars chosen for the study.

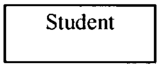
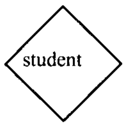
Entity Relationship Diagrams

The ERD was developed by Chen (1976) to provide a graphical method for developing the **data structure** associated with an information system. In the Entity-Relationship (ER) grammar, entities, attributes, and relationships between entities represent the "real" world. For example, in a university environment entities might include students, teachers, and courses. The entity called "student" has attributes such as a number, name, and address. Teachers have attributes like names, titles, and office numbers, while courses have attributes like course name, location, and meeting time. The entities "teacher" and "courses" are related as "Teacher teaches course". In the same way, students and courses are related through "Student takes course". While each individual relation seems simple, the beauty of the ER model is that it is able to use a simple set of constructs to represent a complex real world domain. In this study we are using a basic model for the ERD with an extension for the use of a "Is-a" relation. The "Is-a" relation enables the modeler to create a subclass entity that inherits the attributes from a superclass entity. For example, the subclass entity "Graduate Student" inherits all attributes form the superclass "Student". An example of the ERD that was given to participants in the study is shown in figure 5 below.

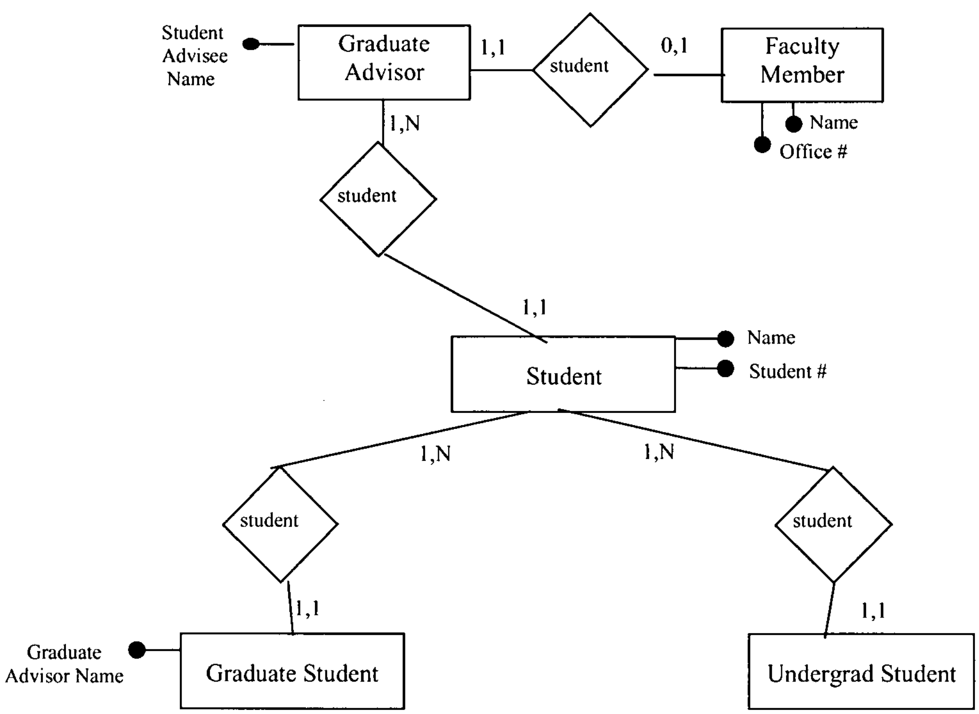
---

[5] The "Is-a" construct is part of the Extended Entity Relationship model (EER) as described in Teorey et.al. (1986) and Storey (1991).

Figure 5: An Example Entity Relationship Diagram

	The rectangular symbol refers to an entity. The name of the entity is provided inside the box
	The diamond shaped symbol refers to a relationship. The name of the relationship is inside the diamond.
	A line with no arrow is used simply to connect entities with relationships
	The knob that is filled in indicates a **mandatory** attribute of an entity. In other words, the solid knob indicates that the attribute **must** be "filled" in and cannot be left blank.
1,N	The (1,1), (1,N), (0,1), and (0,N) symbols indicate cardinalities associated with connections between entities. The **first** number represents the **minimum** number of connections between two entities, the **second** number represents the maximum number of connections. In example below, a graduate student has a minimum of 1 advisor and a maximum of 1 advisor. On the other hand, a graduate advisor has a minimum of 1 student and a maximum of N students that they are advising.

Example:
In this example, all Faculty members have a name and office number and some, but not necessarily all of the faculty are graduate advisors. Graduate advisors can have more than one graduate student as an advisee. All students have a name and student number and there are two types of students. A graduate student must have one and only one graduate advisor. Undergrads do not have a graduate advisor.



Data Flow Diagrams

The DFD, as described by Gane & Sarson (1979), and De Marco (1979) uses a small set of graphic symbols to provide an diagrammatic overview of the ***processes*** in an information system. The DFD has proved to be a popular method as Avison & Fitzgerald (1995, p 162) note:

> *"The data flow diagram (DFD) is fundamental to structured system methodologies and was developed as an integrated part of those methodologies.. However, DFD has been adopted and adapted by a number of other methodologies... Like entity modeling and normalisation, DFDs are an important technique in a variety of system development methodologies."*

DFD's can be created for both physical and logical implementations. The logical DFD describes what a system will do conceptually, whereas the physical DFD describes the technology that is used to accomplish the system objectives. The logical DFD is by far the more popular technique, and from this point forward our discussion will refer only to logical DFD.

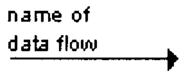
In the DFD grammar, a system is created from four graphic symbols: data flow, process, data store, and external entity. Subsystems can also be identified on the diagram by surrounding related process, data flows, and data stores with dotted lines. The DFD also supports the idea of "levels". The DFD begins with a context diagram, which is highest level abstraction of an information system. Each of the processes in the context diagram can in turn be "exploded" into more detailed and less abstract levels of the system. The concept of leveling enables the developer to partition a complex system so that the system can be more easily understood. The DFD is recognized as a useful analysis technique as Avison & Fitzgerald (1995, p. 163) note:

> *"The graphical aspect means it (DFD) can be used both as a static piece of documentation and as a communication tool, enabling communication at all levels.... The fact that the DFD has proved amenable to users means that it is easier to validate correctness and the possibility of a successful information system resulting is increased. "*

The grammar for the DFD along with an example of the DFD is provided in Figure 6 below. This figure was used as an example for the participants in the study.

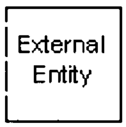
Figure 6: An Example Data Flow Diagram

This symbol with rounded corners indicates a **process**. A process transforms flows of data. A short description of the process is provided in inside the box.

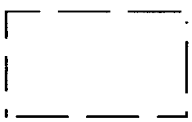


The arrow symbol indicates a **data flow**. Each data flow is identified by a name. The direction of the arrow indicates the direction of the flow of data.

This open-ended rectangle symbol indicates a **datastore**. Data can either be placed into a datastore from a process or data can be retrieved from the datastore by a process.



The square box symbol indicates an external entity. This symbol is normally used to indicate either the source of data or the destination for data.



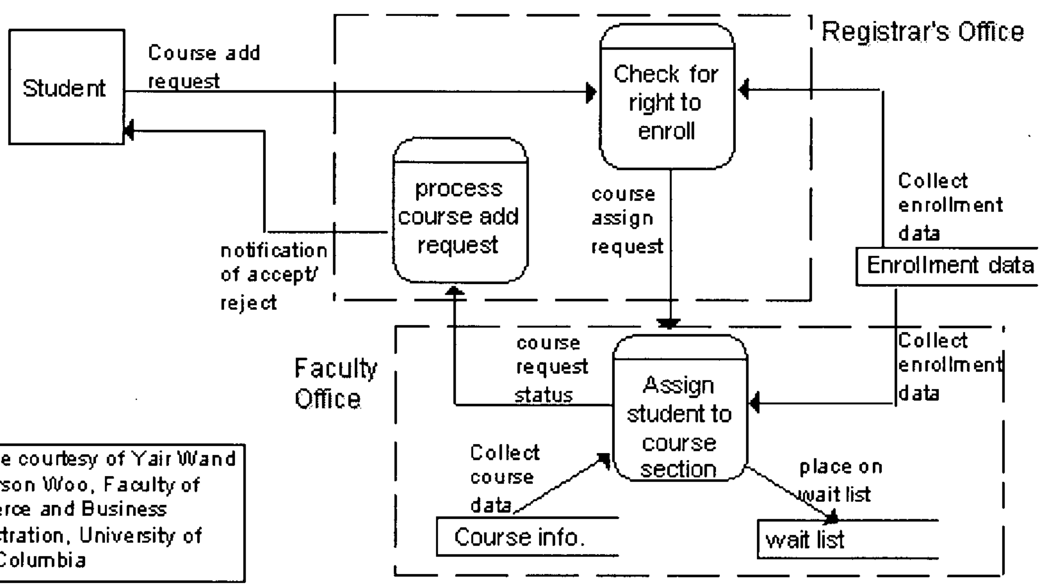
The dashed lines that surround a number of processes, dataflows, and datastores indicate a boundary of a department or important organizational structure.

Example:

Students submit a request to attend a course to the registrar's office. The registrar's office checks if the student has the right to enroll (i.e. proper academic standing, no default on tuition) and then submits the approved requests to the faculties. If a request is rejected, the student is notified.

In the faculty, students are assigned to course sections pending availability. If space is limited, priority is given based on program requirements and credit accumulated so far. The registrar's office is then notified as to requests status and section allocation. Student requests that are not approved due to space limitations are placed on a waiting lists.

The registrar's office notifies students whether their requests have been approved and also notifies students on fees payable based on student status and courses approved.



Example courtesy of Yair Wand and Carson Woo, Faculty of Commerce and Business Administration, University of British Columbia

Combining DFD and ERD

Usually diagramming grammars are not capable of showing all of the potentially important perspectives of an information system. For this reason, system analysis methodologies normally employ several different modeling grammars when describing a system (Green, 1996). Necco, Gordon and Tsai (1987) used questionnaires to discover that many organizations use multiple approaches (data flow diagrams, data dictionaries, flow charts, and text narratives) to develop information systems. Many information system methodologies such as Yourdon's (1989) Modern Structured Analysis, Structured System Development and Design Method (SSADM), or the object-oriented Unified Modeling Language (UML) (Rational, 1999), advocate the use of multiple techniques for representing a system.

One of the natural combinations of analysis grammars is to employ the process elements of a DFD and the data structure elements of an ERD. The DFD and ERD are a popular combination of analysis methods (Green, 1996) because both process and data structures are necessary to understand how a system works. Textbooks in system analysis, such as Hoffer, George, & Valacich (1998), provide examples of how to combine the process oriented DFD with the data oriented ERD. The linkage between the two diagramming techniques is not made directly on the diagrams, but rather has to be understood by the person viewing the two diagrams.

The relationship between ERD and DFD can be described in the following manner. An ERD that is related to a DFD should represent the data structure of all data residing in data store and all of the data incorporated into data flows shown on the DFD. When viewing a DFD, the person viewing the model must make connections between the data stores and data flows on the DFD to the entities and relationships provided in the ERD. Computer Aided Software Engineering (CASE) tools have been designed to make these connections automatically, but these connections can be difficult for many viewers to make. Since the task I have chosen in this study is to represent a conceptual model of a domain, and both **process and data elements** are included in a domain, the combination of DFD and ERD provides a natural second alternative grammar for the study.

Object-Oriented Analysis

The object-oriented approach emerged as a software development discipline that has only recently been adapted to systems analysis (Parsons & Wand, 1997). The adaptation to system analysis has meant the earlier focus on implementation models, that are closer to code, has been replaced with the task of describing a business domain. There are many alternative grammars proposed using object-oriented techniques (Wand, 1989). In this study, the researcher has chosen to use only basic features commonly used in object-oriented analysis grammars. Since there is no standard set of object oriented constructs, the diagramming constructs for the object oriented diagrams used in this study are taken from those provided in Wand & Woo (1993). These constructs are similar to those provided in Coad & Yourdon (1992).

It is understood that simplifying the grammar necessarily reduces the grammar's expressiveness. Simplifying the OOA grammar is important in this study, however, because the simplification enables a greater focus to be placed in the truly novel idea that both data elements and dynamics can be placed on the same diagram. The OOA grammar is interesting because data and behavior are encapsulated in objects. Other important and useful features of the object-oriented approach such as inheritance, association, classification, and polymorphism are left to consider in future studies.

There is no one grammar used for object-oriented analysis. For this reason, some definition of the grammar used in this study is necessary. The description of the object-oriented grammar used in this study is taken from Wand & Woo (1993). The definition has been developed from ontological constructs provided earlier in Table 4. The full development is not presented here. Instead we will look at the definitions that result from the ontological development. The definition provided here begins with the description of objects (Wand & Woo, 1993, p. 4):

> *"...an **object** will be considered a representation of a thing in the modeled domain. The **state** of the thing will be reflected via the **attributes** of the object. The internal transformations will be represented by the actions the object can take. Interaction is represented by external **requests** for **services** sent by objects to other objects...*

From this statement we learn that objects are things in the domain that have attributes, and the value of the attributes represent the state of the object. We also know that objects interact through requests for services. Wand & Woo (1993, p. 7) continue the definition:

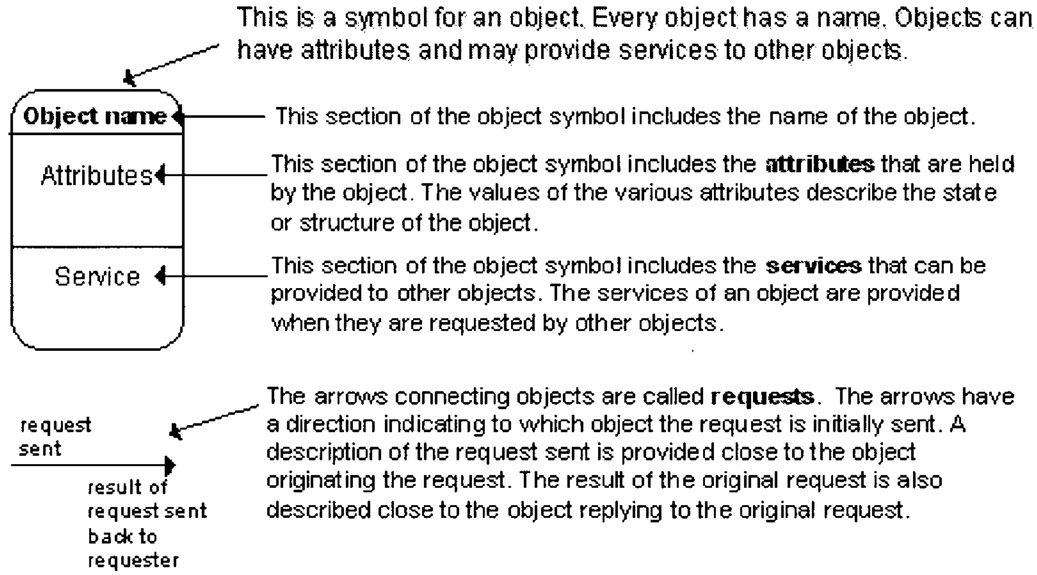
> ***Services** reflect internal transformations of things. These transformations will be invoked as a result of the external events, modeled as **requests** for service. Applying previous rules, and the principle of minimality ... a service will be included in an object if and only if it is invoked directly by at least one request included in the view... such a request may be generated by an external object or by an internal object as part of the systems response to an external request."*

This statement indicates that objects perform services, which are internal transformations of the object that result from a request. The statement also recognizes that services are only included in the diagram if the service is invoked by a request. These elements - object, attribute, requests, and services - make up the basic elements of the object-oriented grammar used in this study. A description of the diagramming constructs is provided below.

The object-oriented grammar used in this study, given the simplified view that has been taken, has few constructs. **Objects** are indicated by rectangles with rounded corners. An object is identified with a name located at the top of the rectangle. Objects have **attributes** that are listed in the top half of the "object" rectangle. As noted earlier, the values of all the attributes of the object, at a point in time, indicate the **state** of the object. Objects also provide **services**. Services are the actions that an object can accomplish. The services are listed in the bottom half of the "object" symbol. An object is said to **encapsulate** both the attributes and services related to an object. The encapsulation of both data and processing elements in a single "object" is the most important difference between object-oriented and structured analysis techniques.

Objects provide their services on the basis of **requests** that come from other objects. A request from one object can result in a request to other objects and is often associated with a reply. In the grammar used in this study, only the **external** attributes and services of the object are shown. External attributes are identified as any attributes that are changed as the result of a request from another object. External services are services that are enacted as the result of a request from another object. An example of an object-oriented diagram is provided below in Figure 7.

Figure 7: An Example Object Oriented Diagram

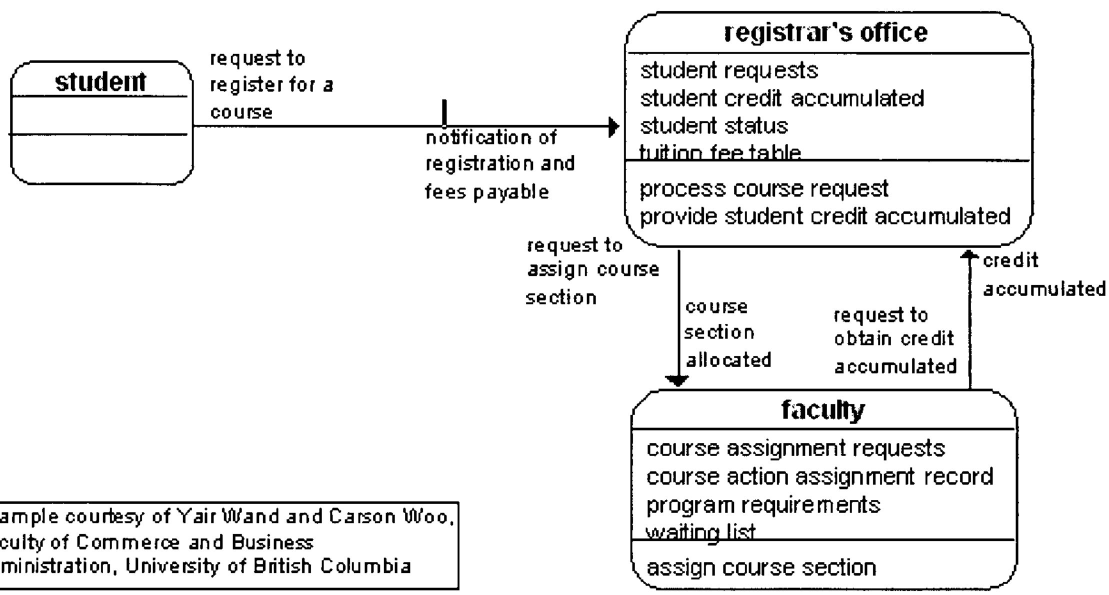


Example:

Students submit a request to attend a course to the registrar's office. The registrar's office checks if the student has the right to enroll (i.e. proper academic standing, no default on tuition) and then submits the approved requests to the faculties. If a request is rejected, the student is notified.

In the faculty, students are assigned to course sections pending availability. If space is limited, priority is given based on program requirements and credit accumulated so far. The registrar's office is then notified as to requests status and section allocation. Student requests that are not approved due to space limitations are placed on a waiting lists.

The registrar's office notifies students whether their requests have been approved and also notifies students on fees payable based on student status and courses approved.



Example courtesy of Yair Wand and Carson Woo, Faculty of Commerce and Business Administration, University of British Columbia

Summary of Methods Used in Intergrammar Study

In the discussion above, the three presentation methods - text description, structured analysis with combined DFD and ER, and OOA – have been presented as alternative techniques for describing a domain. The three methods represent very different approaches to the description of information system domains. These three methods are also representative of the historical movement in system analysis methods in the movement from text based descriptions through structured analysis with separate data and processing elements, and finally towards object-oriented concepts. While the grammars have been simplified and the cases the grammars will be applied to are relatively simple, this comparison has the potential to provide useful information on the relative strengths and weaknesses of the three grammatical approaches. In the next section we develop hypotheses regarding the relative performance of these alternative grammars in communicating information about domains.

4.1.2 Developing Hypotheses for Intergrammar Comparison

Two primary research questions are discussed in this section. The first deals with the expected increase in the effectiveness of communication using diagrams as opposed to textual information. In addressing this question, the text presentation is compared with two diagramming techniques (DFD/ERD and OOA). The second hypothesis addresses the claimed "superior" performance of the object-oriented grammar over the structured analysis techniques of combined DFD and ERD. This hypothesis draws a direct comparison between the two graphical approaches to analysis. For each of the main hypotheses, a variety of sub-hypotheses related to the measures described in Chapter 3 will be formed. A summary of the hypotheses developed in the following sections is provided at the end of the chapter.

Research Question 1: Is a picture worth 10,000 words?

Many researchers take for granted the notion that diagrams of domains are a more effective communication technique than text descriptions of the same domains. For example, De Marco (1979, p 10) notes:

Computer systems analysis is faced with this difficulty (describing systems). Structured Analysis attempts to overcome this difficulty through the use of graphics. When you use a picture instead of text to communicate, you switch mental gears.

> *Instead of using one of the brain's serial processors you use a parallel processor... All of this is a highfalutin way to present a "lowfalutin" and very old idea: A picture is worth a thousand words."*

This common sense notion of the effectiveness of graphic descriptions for system analysis is echoed by Avison & Fitzgerald (1995, p. 163) who write:

> *"The graphical nature of the DFD also means a more concise document, as it is argued that a picture can more quickly convey meaning than more traditional methods, such as textual narrative."*

While the notion that diagrams are a "better" presentation method than text is a appealing, justification for the perceived advantages of graphical models has not yet been articulated, to the best of my knowledge, by researchers in system analysis. This section will discuss some reasons for preferring diagrams.

Before presenting reasons for the preference of diagrams, it is important to contradict the notion that diagrams are "obviously" better than text representations. Two important assumptions underlying the "obvious" appeal of graphics should be considered. First, the notion that graphics outperform text descriptions assumes that the graphics provide useful and usable information to the person viewing the model. This remains an empirical issue as our limited knowledge of cognition does not provide a theoretical basis for the argument. This issue will also depend on the grammar associated with the graph. Unfortunately, I am aware of no studies that compare text representations to graphics in system analysis. The second assumption is that individuals have a "natural" ability to understand graphic information. More importantly, this natural ability outweighs the years of training and experience that most individuals have undergone in developing reading and writing skills. It could be argued, for example, that our experience with text has enabled individuals to develop sophisticated cognitive representations from text descriptions without the need for diagrams. If this were true, why would we need graphic descriptions? Finally, it should also be noted that text is unquestionably a more expressive grammar than any of the graphic grammars developed to date, for analysis. When we represent a domain graphically, we are necessarily trading away expressiveness in the hope for improved clarity. It is not obvious, to this researcher, that this is trade-off is always successful.

We have noted that many researchers believe diagrams are "better" methods for presenting information than text descriptions. To understand if diagrams are better than text, it is necessary to understand what the text and diagrams are supposed to do. In this study, we have identified the objective of analysis methods as the development of understanding in an individual viewing the description of a domain. A "good" presentation method will assist a person in the process of interpretation, that is, viewing the domain description in order to develop an understanding of how the system works. The question is, what features of diagrams enable diagrams to outperform text descriptions in the interpretation process?

Mayer & Gallini (1990, p. 715) provide two suggested benefits of diagrams (illustrations):

> *"Based on theories of mental models ... we identify two major features of illustrations that could help learners build runnable mental models: **system topology** and **component behavior**. System topology refers to the portrayal of each major component within the structure of the system.... Component behavior refers to the portrayal of each major state that each component can be in and the relation between state change in one component and state change in other components. "*

Mayer & Gallini suggest, therefore, that when diagrams portray major system components along with their behaviour within the system, the diagrams provide information beyond that provided by text descriptions alone.

The advantage of diagrams is that they immediately convey a structure of a domain by selecting important items from the domain and drawing relations between these items. Diagrams are "considerate" descriptions because diagrams select and organize important information from the domain. Diagrams also can show a direct linkage between items in a domain - for example drawing a line between two items on a diagram - whereas text can only make references to the linkages. Relations made visually would be expected to outperform relations made through text as the visual relation provides more organization of the domain information. Diagrams are also necessarily less verbose than text descriptions. This suggests that diagrams help individuals in understanding by reducing the number of items that an individual has to assimilate into memory. For these reasons, graphical models of a domain are expected to outperform text descriptions for the interpretation process.

Before we develop the predictions related to the comparison of diagrams and text descriptions, it is important to note that what has been stated above regarding the comparison or text and diagrams is largely conjecture and not based on an underlying theory of information processing. To the best of this researcher's knowledge, no theoretical comparison of understanding developed from text descriptions and diagramming methods has been undertaken. This is not surprising, due to the very different nature of the presentation methods. The hypotheses and hypotheses that follow in this section, therefore, should be viewed as exploratory working proposition and predictions. The fact that the predictions in this section are exploratory does not, however, diminish the value of the data collected from the study, nor does it negate the findings. The lack of theory restricts our ability to infer why effects were observed. Having warned the reader, the following hypothesis is developed:

Hypothesis 1

H1.1: Diagrams will be better able to promote understanding in individuals interpreting a domain description than text descriptions.

Diagrams should promote improved levels of understanding, as argued above, because diagrams are better able to convey system topology and component behavior of an information systems domain than text descriptions. No attempt will be made in this thesis to separate the effects of system topology or component behaviour, so it will be impossible to discern which of these effects leads to observed differences. Of interest for this researcher, is not the separate effects of system topology and component behaviour, but rather, the observation that there is difference between individuals viewing diagrams or text.

Referring to the empirical foundations described in Chapter 3 and the discussion above, we can make several predictions regarding the relative performance of diagrams over text descriptions

Predictions

1. ***The scores on comprehension questions will be higher for graphical presentation methods than text presentation methods, particularly for questions that require knowledge of the relationship between two items in a domain***. Justification for this prediction is based on the superior selection and organizing features provided in the diagrams, particularly for relations between items in a domain. This prediction recognizes that even though text description can be more expressive, there is no significant effect on the level of comprehension attained through the text description.

2. ***Problem solving performance will be higher in individuals that have been provided with a graphical model than for those participants who are shown a text description***. It is assumed that the superior selection and organizing processes provided by diagrams will help to encode a more sophisticated cognitive model of the domain. This model will lead to more creative problem solving in individuals provided with diagrams as opposed to those provided with text descriptions.

3. ***Performance on the Cloze test will be higher for individuals that have been provided with a test description than for those participants who are shown a diagram.*** The reason for this statement is that the Cloze test is created from the original text description. Participants provided with text descriptions can simply use memory recall to fill out the Cloze test. This recall is in addition to the information they have learned and understood from the study. We would expect, therefore, with this additional advantage that the Cloze test results for the group using text representation to be significantly higher than those using diagrams.

4. ***Individuals who are provided with text representations will take more time to complete the comprehension, problem solving and Cloze test that persons provided with a diagram.*** The increased time associated with developing an understanding text descriptions is directly related to the cognitive effort required to create the system topology and component behaviour that are more apparent in diagrams.

5. ***Individuals who are provided with diagrams will perceive them as easier to use than text descriptions.*** This prediction is based on the common sense notion that graphics are easier to use than text. Of course, the participants are much more familiar and comfortable with text descriptions than diagrams, so the perceived ease of use may be affected by this familiarity.

In the next chapter, empirical methods are described to collect data for assessing the first hypothesis. We can now turn our attention to the second hypothesis in this intergrammar study.

Research Question 2: Structured Analysis vs. Object-Oriented Methods

Proponents of object-oriented analysis (OOA) methods claim OOA methods are superior to more traditional methods such as DFD or ERD for a variety of reasons. For instance, Jacobson (1995, pp. 68-9) states that OOA models are more "comprehensible", "understandable", "changeable", "adaptable", and "reusable" than other methods. Coad & Yourdon (1991, pp. 3-4) suggest that OOA enables analysts to tackle more complex problems, to improve analyst/user interaction, to increase internal consistency, to provide more explicit representation of commonality, and to develop a more consistent underlying representation. Booch (1991, pp. 77-8) suggests OOA has further advantages such as stability in design, reduced risk in developing complex systems, and a greater proximity to the workings of human cognition.

The claimed advantages of OOA have, therefore, been substantial. There is surprisingly little empirical evidence, however, for superior performance of OOA over other analysis methods. Two separate empirical studies (Vessey & Conger, 1994; Wand, Gemino, & Woo, 1997) have indicated a preference, among novice analysts, for process oriented methods such as DFD's. This preference is described clearly by Vessey & Conger (1994, pp. 111) who report:

> "*All the results presented here suggest that novice analysts found the process methodology easier to use than the data methodology, which in turn, was easier to use than the object methodology... The results demonstrate that, of the three methodologies investigated, novice analysts were better able to apply the process methodology (DFD) and the least able to apply the object methodology.*"

These results seem to contradict some of the claims, noted previously, made by proponents of object oriented methods, and suggest a research question: **Are there interpretation advantages in adopting OOA analysis methods over alternative analysis methods such as DFD & ERD?** This question motivates the comparison of the OOA grammar with the combined DFD/ERD grammars.

It would be useful, at this point, to utilize an ontological analysis of an OOA grammar and the combined DFD/ERD grammars. Unfortunately, no such analysis exists. And the development of these ontological comparisons requires resources beyond the reach of this researcher. Future theoretical analysis of structured analysis and object oriented analysis will provide a much needed theoretical basis for this comparison[6].

In the absence of an ontological comparison, there are two possible sources for the differences between DFD/ERD and OOA. The first comes from the observation that the OOA captures data and process elements in a single diagram, whereas the DFD/ERD combination requires two diagrams. The use of two diagrams requires the person viewing the models to actively create the linkages between the two diagrams as noted by Avison & Fitzgerald (1995, p. 271). Describing the creation of these linkages Gane & Sarson (1979) note:

> *"The physical files on a database will then be designed. They will be based on the data store contents previously specified at the logical level. Data stores are defined in the DFD as the temporary storage of data needed for the process under consideration. This has the effect of introducing many data stores scattered all over the DFD. Many of these will be similar in content and have a significant degree of overlap. The data stores need rationalising, and the technique of normalisation... is utilized to simplify the data stores into logical groupings. The actual process mapping and the design of the physical files (or databases) are not defined..."*

[6] An informal analysis of data flows grammar by Yair Wand and Ron Weber has indicated that the DFD does not fit easily into the set of ontological constructs proposed in the BWW model. On the other hand, the object oriented concepts provide a good fit with the BWW ontological constructs. This report is only informal and a more detailed analysis is required before conclusions regarding the DFD and OOA can be reached.

Linking the information provided in the two diagrams can be a difficult task for system analysts. For a person with little or no experience with the DFD and ER grammars, the two diagrams may be treated as completely independent. Whatever the case, the OOA may be a preferable grammar because it does not require linkages to other diagrams using different grammars.

A second reason for the advantage of OOA grammar is the purported "natural" fit with the way that individuals think about the world. Although these claims are often made by proponents of object oriented methods, the researcher has been unable to uncover a theoretical basis for these claims. Proponents of object oriented methods often use an example to highlight the advantages of objects. Take for example a common object - a dog. When we think about a dog we think of things like color of the hair, the breed of the dog, the size of the dog, the bark, does the dog sit on request, will the dog roll over, and many other things. Notice that we do not naturally separate the dog's attributes (color of hair, size, and breed) from the dog's behaviour or "services" (barking, sitting, and rolling over). Perhaps the separation of systems into data elements and process elements, while logical, is not a natural separation that individuals make. The OOA grammar may be closer to the way in which individuals think about the "real" world. If this is true then the OOA may enjoy an advantage in promoting understanding in individuals interpreting a diagram because the objects on the diagram are closer to the type of "cognitive objects" that individuals use to think about the world around them.

It is again important to note, that in the absence of theory regarding the comparisons made between the combined DFD/ERD and OOA diagrams, the hypotheses and predictions noted in this section should be viewed as exploratory. As noted earlier, however, the fact that the predictions in this section are exploratory does not diminish the observed effect or negate the findings. The lack of theory restricts our ability to infer why effects were observed. This suggests that future comparisons will benefit from a theoretical basis upon which to build hypothesis regarding comparative performance of the alternative grammars.

The two arguments developed above regarding the potential advantage of OOA over DFD/ERD grammar enable us to state a second hypothesis:

Hypothesis H1.2

H1.2: The OOA grammar will be better able to promote understanding in individuals interpreting a domain description than the combined DFD/ERD grammar

This hypothesis is made based on the argument that the OOA grammar will out perform the combined DFD/ERD grammar for two reasons. First the OOA grammar constructs are closer to individual cognitive constructs and, second, the OOA models do not require linkages to be developed between two separate diagrams. Since the argument is based on two factors, we will not be able to separate the individual contribution from the two factors.

Referring to the empirical foundations described in Chapter 3 and the arguments made above, we can make predictions regarding the relative performance of diagrams over text descriptions. These predictions are provided below:

Predictions

1. **Comprehension scores for individuals provided with DFD/ERD diagrams will not be significantly higher than individuals provided with an OOA diagram.** This prediction is based on the fact that the information for answering comprehension questions is provided on both the OOA and DFD/ERD diagrams. Since the information is available on both diagrams, no difference is expected.

2. **Problem solving scores for individuals using OOA will be significantly higher than individuals using the DFD/ERD diagrams.** This prediction follows from the argument that the OOA constructs are closer to "natural" cognitive constructs, and OOA diagrams are complete in that they do not require linkages to other diagrams. The OOA constructs would then promote more sophisticated cognitive models that would promote more understanding and enable participants to develop more creative solutions to problem solving questions. 0

3. **Cloze scores for the participants provided with OOA will be higher than participants provided with DFD/ERD.** Since neither DFD/ERD participants nor OOA participants will have seen the original text description, the Cloze test in this case acts like a test of the overall understanding of the domain. The argument that OOA will promote a higher level of understanding than the DFD/ERD grammar suggests that Cloze test scores should be higher for participants provided with OOA descriptions.

4. **The time taken to complete the comprehension, problem solving, and Cloze test will be shorter for participants using OOA.** This prediction again relies on the argument that the linkages required between the DFD and ER diagrams will require extra processing time. Also the "natural" OOA constructs should be understood more quickly than DFD/ERD constructs thereby reducing the time necessary to complete the tasks.

5. **Perceived ease of use scores will be higher for OOA participants.** The argument that OOA constructs are similar to cognitive constructs suggests that individuals will find the OOA grammar easier to use than the DFD/ERD grammar.

The predictions noted above will be revisited in Chapter 7 where the results from the empirical study are presented. The hypotheses in this section have been concerned with intergrammar comparison, that is, the comparison between alternative grammatical approaches. In the section that follows, our attention will be narrowed to a single grammar defined in the entity relationship model. Alternative grammars for developing the ER model are presented and hypotheses for the intragrammar comparisons are developed.

4.2 Study 2: Intragrammar Comparison

The second study described in this thesis uses the instrument described in Chapter 3 to focus on an intragrammar comparison. Since the differences in the grammars are relatively small, this second study provides an excellent opportunity for testing the sensitivity of the newly adapted empirical instrument. In addition, since the ontological analysis for the Entity Relationship Diagrams has been completed, a theoretical difference between the two grammars can be created and assessed. The comparison is sparked by a specific research question raised by Bodart & Weber (1996). These researchers state, in the context of Entity Relationship Model (ERM), that a grammar that uses subtypes with mandatory ("must have") properties is ontologically superior to a grammar that uses optional ("may have") properties. On the basis of this argument, it is hypothesized that individuals should develop superior cognitive models using the sub-typing approach. The superior cognitive models, in turn, would lead to a significantly better score on a free-recall task that asked participants to "re-draw" a representation after viewing it for three minutes.

Unfortunately, while the direction of the effect in Bodart & Weber was as anticipated, the differences could not be considered significant. A revisiting of Bodart & Weber is suggested for the second study in this thesis. The same experimental case as was used in Bodart & Weber (1996) will be used. The instrument described previously in Chapter 3 that makes use of three dependent measures of performance (comprehension, problem solving, and recall) will be used.

4.2.1 Introduction to Grammars for the ERD

The entity relationship model (ERM) and the related entity relationship diagram (ERD) developed by Chen (1976) "*...has proved very useful and is included in many information* *system development methodologies.*" (Avison & Fitzgerald; 1995, p. 148). An example of a ERD was provided in Figure 5 above. In this study, we will be using the ERD at a conceptual level, to describe a domain. The original ER diagramming technique has undergone a series of extensions that have been proposed by a variety of researchers and practitioners. These extensions have resulted in the development of a number of alternative techniques that utilize the core ERM constructs. These techniques differ either in the symbols used to construct ERD or in the definition of the constructs used for modeling data structures. Since the methods use different symbols and definitions, at times these alternative grammars for the ERD can conflict in the way that they "handle" domain information.

This study is concerned with two particular grammars used in creating an ERD. To illustrate the differences between these grammars we will use an example from Wand, Storey, and Weber (1993, p. 24). The example is summarized below:

> **Example: Students and Faculty Advisor**
> *In this example, there are students and faculty members. A student may be a graduate student or an undergraduate student. Graduate students have one and only one graduate advisor. A graduate advisor must be a faculty member, but a faculty member does not have to be a graduate advisor.*

The "Mandatory" Grammar

There are many ways of describing the information provided above using an ERD. The models in Figure 8 highlight the differences between two particular ERD grammars: the mandatory and optional approaches. The model using mandatory attributes and subtyping of entities requires that more general entities, such as student and faculty member, be split into subtypes with mandatory properties. A subtype is indicated using an arrow symbol with the arrow pointing towards the more general class. This is described in the top section of Figure 8. The general entity "student" is then represented as an entity with two subtypes: graduate student and undergraduate student. Note the arrows pointing from "graduate student" and "undergraduate student" towards the more general "student" entity. In a similar way, the subtype "graduate advisor" is created from the entity "faculty member."

All relationships in the subtyping model are mandatory so that the minimum cardinality requirement that can be used is one. All attributes are "mandatory" so that all of the attributes relate to the particular subtype. Mandatory attributes are shown as "filled-in pegs" in the diagram. In requiring mandatory properties and relationships, the mandatory with subtyping model provides a well defined, but somewhat complex, representation of the situation. The ERD developed using the 'mandatory" grammar is provided at the bottom left of Figure 8.

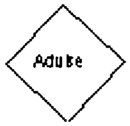
The reader should make a careful distinction between the use of conceptual models and the use of data models for record structure design. The term "mandatory", as used in conceptual design described above, does not imply that all underlying fields in a resulting record structure must be non-null. Assigning mandatory attributes to an entity on a conceptual ERD only indicates that the attribute (and resulting field) must be attached to the entity it helps to describe. The fields in an underlying structure may remain null if information is not available, regardless of whether the field is "mandatory" or "optional". Conceptually, the mandatory property indicates only that a field that "must" be a part of the entity's record structure, not that a field must be non-null. This notion will be discussed further at the end of this section.

The "Optional" Grammar

Models using optional attributes, in contrast to mandatory grammar, allow for both optional attributes and optional relationships in the ERD. Optional attributes occur where a faculty member "may be" a graduate advisor. An optional attribute is indicated by "pegs" attached to the entity that are not filled in (they are optional). Note that while optional attributes can be used, mandatory attributes can also be attached to entities. Optional relationships occur where students "may have" a graduate advisor. In an optional relationship the minimum cardinality requirement is "0". Note the "(0,1)" and "(0,N)" cardinalities attached to the relationships in the right hand side of Figure 8

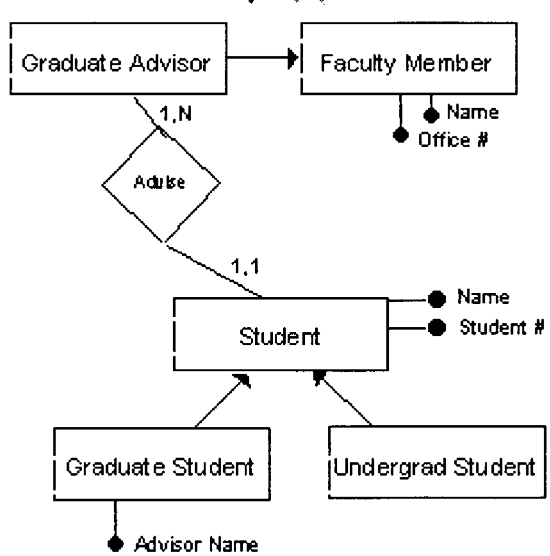
As can be seen in Figure 8, the diagram created using optional properties is initially less complex, in that it has less symbols than the mandatory model. The simplicity may come at a cost, however, since the optional diagram is necessarily more ambiguous than the mandatory model. To correctly interpret the optional diagram, a person viewing the model would have to rely on personal experience to reduce he ambiguity introduced by optional components.

Figure 8: Using Subtypes with Mandatory Properties or Optional Properties

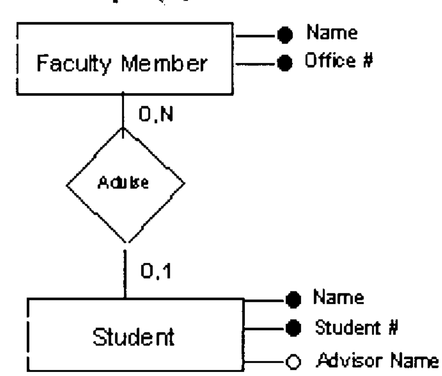
Student	The rectangular symbol refers to an entity. The name of the entity is provided inside the box
	The diamond shaped symbol refers to a relationship. The name of the relationship is inside the diamond.
———	A line with no arrow is used simply to connect entities with relationships
———▶	A line with an arrow is used to point from an entity subtype to a more general entity. In example (A) below, the line with an arrow indicates that a graduate student is a type of student (graduate student is a subtype of the more general entity called student). Note that lines with arrows are only used to point from one entity to another more general entity.
—● Name	The knob that is filled in indicates a mandatory attribute of an entity. In other words, the solid knob indicates that the attribute must be "filled" in and cannot be left blank.
—○ Advisor	The knob that is not filled in indicates an optional attribute of an entity. In other words, the knob that is not filled in indicates that the attribute does not have to be given a value. However, a value other than null can be placed in the attribute as required
1,N	The (1,1), (1,N), (0,1), and (0,N) symbols indicate cardinalities associated with connections between entities. The first number represents the minimum number of connections between two entities, the second number represents the maximum number of connections. In example (A) below, a graduate student has a minimum of 1 advisor and a maximum of 1 advisor. On the other hand, a graduate advisor has a minimum of 1 student and a maximum of N students that they are advising.

Examples:
The examples (A) and (B) below both depict the same situation. In these examples, all Faculty members have a name and office number and some of them are graduate advisors. All students have a name and student number. Graduate advisors can have more than one graduate student as an advisee. A graduate student, however, can only have one graduate advisor. Faculty members are not advisors for undergrads.

Example (A)



Example (B)



The optional and mandatory models differ in their appearance, yet both models in Figure 8 can be shown to be "semantically equivalent" (Weber and Zhang, 1996). Semantic equivalence implies that each of the models could be used, further on in the development process, to independently implement **exactly** the same relational schema or database design. The fact that the two models can represent the same information does not imply, however, that separate persons viewing the two models will interpret the models in the same way. The two diagrams may differ, therefore, in their ability to promote understanding of the domain in persons viewing the diagrams. An effort has been made in this study to develop an empirical comparison of the understanding developed by persons viewing models created from these competing ERD grammars.

4.2.2 An Ontological Analysis of Alternative Grammars

The assertion that grammars using subtypes with mandatory properties will produce more understandable models than grammars that allow optional properties rests on the BWW (Bunge-Wand-Weber) ontological[7] model proposed by Wand and Weber (1990, 1993) and developed further in Wand, Storey, and Weber (1998). As noted in Chapter 3, the BWW ontology is intended to provide a set of constructs that are capable of representing real-world phenomena related to the development of an information system.

Wand, Weber, & Storey (1993) argued that the use of optional properties in conceptual ERD grammars suffers from the ambiguity (construct overload) associated with collapsing two schema definitions into a single entity. An example will help to clarify this argument. Take the example of a telephone company. The company collects information about customers. Some customers have a pager number and some do not. Initially the pager number may be viewed as an "optional" attribute (property) and modeled as an optional attribute associated with the entity customer. From the telephone company's perspective, however, a customer with a pager number is conceptually different from a customer with a pager number. Customers with pager numbers have at least two account numbers and different options associated with the pager number. The person modeling this domain faces the choice of either splitting the entity customer into two subtypes (Regular Customer, Pager Customer) each with their own "mandatory" attributes, or collapsing "Regular" and "Pager" customers together into one "Customer" entity. The collapsed Customer entity will need to include "optional" properties that are associated with "Pager" customers but need not be associated with Regular customers. Wand, Weber, & Storey (1993) and Bodart & Weber (1996) argue that using mandatory attributes, as compared to optional properties, will reduce the ambiguity in schema definition and lead to a more understandable model. The intergrammar study proposed in this chapter is designed to address this question.

Bodart & Weber's (1996) analysis of optional properties suggests that at a more detailed level of analysis the optional property stands for the ontological construct of a *negated property*, that is a property that a thing does ***not*** possess. They also argue, referring to Bunge (1977, p. 60), that negated properties have no ontic correlate because people do not think about the world in that way. They argue that humans do not conceive of a person or object in terms of the properties that they do ***not*** have. For example, when we think about a dog, we do not generally consider the properties that a dog does *not* have; for example, wings, webbed feet, or walking upright on two legs. Instead we think of the properties the dog has; such as height, breed, and color of coat.

The lack of an ontic correlate for optional properties implies that a grammar that includes the optional property construct cannot be directly related to one of the basic ontological constructs suggested in Wand & Weber (1990,1993). The presence of an optional property in the ERD grammar, therefore, introduces ambiguity in the mapping from grammatical constructs to ontological constructs. This ambiguity introduced by including optional properties reduces what Wand and Weber have defined as the *ontological clarity* of the grammar. The result is that models developed using optional properties, other things being equal, will be theoretically more difficult to understand than a model without the optional property. The theoretical difficulty in understanding is directly related to the reduced ontological clarity associated with the optional property construct.

The argument for preferring models without optional properties to models that allow optional properties can now be outlined. Weber & Zhang (1996) suggest that grammars using optional properties can be represented using grammars that do ***not*** use optional properties but instead use subtypes with mandatory properties. It follows from this that grammars with optional properties cannot be more *ontologically complete* than grammars using subtypes with

[7] Wand and Weber (1993, pp. 220) define ontology as "the branch of philosophy concerned with articulating the nature and structure of the world."

mandatory properties (other things considered equal). The previous discussion has also suggested that grammars using optional properties have less *ontological clarity* than grammars using subtypes with mandatory properties. From these two statements we can conclude that grammars using optional properties have at best the same ontological completeness and necessarily less ontological clarity than grammars using subtypes with mandatory properties. If we accept the hypothesis that grammars that are more ontologically clear, and no less ontologically complete, are more understandable, then it can be suggested that models created from grammars using subtypes with mandatory properties will theoretically be more understandable than models created from grammars using optional properties.

4.2.3 Hypothesis for Study 2: Optional vs. Mandatory

Before describing the hypothesis related to this study, it is necessary to understand why we are revisiting this question. Remember that Bodart & Weber maintained that participants using the mandatory grammar would outperform participants using the optional grammar. Bodart & Weber 91996) used comprehension and verbatim recall to compare understanding across two treatment groups, Bodart & Weber (1996) found little difference in comprehension scores across treatment groups and a slight opposite effect in verbatim recall. Both of these results are in the direction that Mayer (1989) would predict. Individuals who develop complex conceptual models, perform less well on verbatim recall, and generally are no different on comprehension. The difference between Bodart & Weber and this study is the problem solving instrument

Given this background, this study is concerned with the empirical testing of the following hypothesis:

Hypothesis P2.1

P2.1: ERDs that use a grammar including subtypes with mandatory attributes will promote a higher level of understanding in persons viewing the model than ERDs developed using a grammar allowing the use of optional properties

The higher performance associated with the mandatory grammar occurs because grammars with mandatory attributes are more ontologically clear and no less ontologically complete than grammars using optional properties. The added ontological clarity associated with the mandatory grammar should provide an increased opportunity for development of an improved level of understanding.

A theoretical argument justifying this hypothesis has been made above. On the basis of this argument, predictions regarding the relative performance of the "mandatory" and "optional" grammars can be made. The predictions have again been divided into two groups. The first three predictions made regarding the dependent measures are directed towards the relative effectiveness of grammars in promoting understanding in persons viewing the model. The prediction related to perceived ease of use and times required to complete tasks are related to how efficiently the grammars promote this understanding. We will refer to five predictions:

Predictions

1. **Comprehension scores for individuals provided with ERDs created using an "optional" grammar will NOT be significantly higher than individuals provided with ERDs created with a "mandatory " grammar**. Given the relatively small differences between the constructs in the two grammars, the differences in conceptual information presented in the diagrams is likely to be small. This suggests that the comprehension scores across the optional and mandatory groups will not differ significantly This result is likely for most intragrammar comparisons.

2. **Problem solving scores for individuals provided with ERDs created with a "mandatory" grammar will be significantly higher than individuals provided with ERDs created with a "optional" grammar**. This prediction follows from the argument that the mandatory grammar produces a diagram with a higher ontological clarity than the optional grammars. This ontological clarity reduces the ambiguity associated with the ERD. The optional grammar, being less ontologically clear, introduces ambiguity into the ERD and will result in less well-formed cognitive models of the system domain. The higher ontological clarity associated with the mandatory grammar can promote, therefore, a higher level of understanding, which will result in higher scores on the problem solving test.

3. **Cloze scores for the participants provided with ERD created using the mandatory grammar will be higher than participants provided with ERDs crated using the optional grammar.** Since neither "mandatory" participants nor "optional" participants will have seen the original text description, the Cloze test can act as a test of the overall understanding of the domain. The argument that the mandatory grammar will promote a "better" cognitive model of the domain than the optional grammar suggests that Cloze test scores should be higher for participants provided with mandatory descriptions.

4. **The time taken to complete the comprehension test and Cloze test will be shorter for participants using ERDs created suing the mandatory grammar**. This prediction again relies on the argument that the ontological clarity associated with the mandatory grammar will reduce the ambiguity introduced during the interpretation process. The reduced ambiguity in the mandatory grammar. will result in less time sorting out the ambiguities and hence less time required to answer comprehension questions and fill in the Cloze test.

5. **No difference in the perceived ease of use scores will be observed between participants provided with ERD's created from either the mandatory or optional grammar**. This prediction follows from the argument made in Prediction 1 above. The similarity between the ERD's created by the two methods suggests that individuals would find little difference in the perceived ease of use between the two methods.

The hypothesis and related predictions noted above will be revisited in Chapter 7 where the results from the empirical study are presented. The hypotheses in this section have been concerned with intragrammar comparison, that is, the comparison between grammatical approaches within the same grammatical approach (ERD). In the section that follows, our attention is focused across both Studies 1 and 2 to discuss the effect that learner characteristics – prior knowledge of the domain or presentation grammar – can have on promoting understanding.

4.3 Effects of Prior Knowledge on Understanding

The overview provided in Figure 3 in Chapter 3 indicates two important characteristics of the "learners" that have not been considered in the hypotheses and prediction developed in the two studies above. The important characteristics are "prior knowledge of the domain" and "prior knowledge of the presentation method." There are good reasons why the effects of prior knowledge have been ignored in developing the hypotheses for the two studies. Randomization of the participants between groups should eliminate the systematic effects of prior knowledge of the domain and/or the presentation method. This does not mean that the effects of prior knowledge should be ignored. The discussion of empirical methods in Chapter 5 will outline methods for collecting pretest information regarding participant's prior knowledge of the domain and presentation methods used in the study. This section discusses how this type of pretest information could be used to develop hypotheses regarding the effect of prior knowledge on the promotion of understanding.

4.3.1 Focus on Prior Knowledge of Presentation Methods

The first thing to point our regarding the use of prior knowledge is the fact that the prior knowledge of the domain being described in the study is difficult to measure. Two measures are used in the pretest, as described in Chapter 5 and developed in Mayer (1989), to asses the level of prior knowledge about the domain. Participants are first given a short description of the domain and then asked to rate their perception of the knowledge they have about that domain on a seven-point scale. Second, participants are provided with five activities related to the domain of interest and then asked to indicate which of the activities they have participated in. These two measures provide a useful "first cut" measure of prior knowledge about a domain; a measure good enough to assess whether prior knowledge has been randomized across groups. These two measures have **not** been designed to be sensitive enough, however, to provide an adequate assessment of prior domain knowledge that could be used for hypothesis testing. For this reason, the results regarding the effects of prior domain knowledge should be viewed with some skepticism.

Prior knowledge of the presentation methods was also collected in the pretest. Unfortunately, no verified measure of 'expertise" in system analysis have been developed. In the absence of an accepted method for determining the expertise with analysis methods, we developed three measures of prior understanding of system analysis grammars. The first measure asks the participant whether they have used the presentation methods (analysis techniques). The second measure asks participants to indicate their experience (how many months) with the methods. Finally participants are asked to rate their perceived familiarity, confidence, and competence in using the presentation methods. These three measures, as is argued in Chapter 6, provide a useful measure of the prior knowledge of methods. The existence of a measure of prior knowledge of methods enables hypotheses regarding the ability of prior knowledge of a presentation method to promote understanding to be tested. A hypothesis and related predictions are developed in the section below.

4.3.2 Hypotheses of the Effect of Knowledge of Method on Understanding

The argument for the effect of prior method knowledge on the level of understanding is relatively straightforward. The more experience and competence a person has with a presentation method, all other things equal, the better the learning outcome when the presentation method is used. Knowledge of the grammar underlying the presentation method should improve both organization and integration of information presented. Improving the organization and integration should lead to a "better" cognitive model and promote a higher level of understanding . We would, therefore, anticipate the following hypotheses:

Hypothesis P3.1

P3.1 Individuals with more experience and competence with a presentation method used to represent a domain, will develop higher levels of understanding when compared with individuals who have less experience and competence with a presentation method.

This argument has been developed above. On the basis of this argument, predictions regarding the relative performance of individuals with and without prior knowledge of methods can be made. The predictions have again been divided into two groups. The first three predictions are directed towards the relative effectiveness of prior method knowledge in promoting understanding in persons viewing the model. The remaining predictions are related to how efficiently the grammars promote this understanding as measured by perceived ease of use and time to complete tasks.

Predictions

1. **Comprehension scores for individuals who have prior knowledge of the presentation method will be higher than comprehension scores from individuals who have less prior knowledge of the presentation method.** This prediction is based on the assumption that experience with the presentation method provides a cognitive advantage over those individuals with little or no experience with the presentation. The advantages may come from improved integration of information or from the organization of information that the representation provides.

2. **Problem solving scores for individuals who have prior knowledge of the presentation method will be higher than problem solving scores from individuals who have less prior knowledge of the presentation method.** Prior knowledge of a method should enable individuals to develop more sophisticated cognitive models of a domain. These models should provide participants with prior competence with presentation methods to develop a higher level of understanding and hence a higher score in the problem solving task.

3. **Cloze test scores for individuals who have prior knowledge of the presentation methods will be higher than Cloze test scores from individuals who less prior knowledge of the presentation methods.** Individuals - with both high and low levels of prior knowledge of methods - who have not seen the original text description can be used to test this prediction. The Cloze test in this case acts as a test of the overall understanding of the domain. The argument made earlier regarding the improved cognitive model associated with high level of prior knowledge about a method suggests that Cloze test scores should be higher for participants with previous method experience.

4. **Time taken to complete the comprehension and Cloze tasks will be higher for those individuals with less prior knowledge with the presentation methods.** This prediction follows the arguments made above regarding the improve ability to integrate and organize information on the representation.

5. **Ease of Use scores for individuals who have prior knowledge of presentation methods will be higher than ease of use scores from individuals who do NOT have prior knowledge of the presentation method.** Knowledge of the presentation method should make the perceived learning process easier as less cognitive effort is required when the grammar underlying the presentation method is understood.

The hypothesis and related predictions noted above are revisited in Chapter 7 where the results from the empirical study are presented. A summary of all of the hypotheses and predictions made in this chapter is provided below in Table 3.

A few notes regarding Table 3 are required. The hypotheses are listed along the top of the table. There are four hypotheses, related to the four hypotheses developed in this Chapter. Each row in the table indicates either a measure of the product of understanding (comprehension, problem solving, Cloze), or a measure of the interpretation process (time to complete task, perceived ease of use). The hypothesis for the time taken in each of the three tasks is provided in the row labeled “time to complete task”. The predictions listed in the discussion above are placed in the cells of the table using the following symbols: “>” (greater than); => (greater than or equal to); “<” (less than); “=<” (less than or equal to); “=” (equal to); “<>” (not equal to).

Table 3: Summary of Hypotheses and Predictions

		Hypotheses					
	Type	**Measure**		**H1.1**	**H1.2**	**H2.1**	**H3.1**
Predictions	Product	Comprehension		Graphic> text	OOA => DFD/ERD	Mand = Opt	HPKM* > LPKM**
		Problem Solving		Graphic > text	OOA > DFD/ERD	Mand > Opt	HPKM > LPKM
		Text Reconstruction (Cloze test)		Graphic < text	OOA > DFD/ERD	Mand > Opt	HPKM > LPKM
	Process	Time to Complete Task	Comp.	Diagram < text	OOA =< DFD/ERD	Mand =< Opt	HPKM < LPKM
			Prob.	Diagram < text	OOA =< DFD/ERD	Mand =< Opt	HPKM < LPKM
			Cloze	Diagram < text	OOA =< DFD/ERD	Mand =< Opt	HPKM < LPKM
		Perceived Ease of Use		Graph > text	OOA => DFD/ERD	Mand = Opt	HPKM > LPKM

Abbreviations used in Table 3

Graph	- graphical description	Text	- Text description
OOA	- Object Oriented Analysis	DFD/ERD	– Data flow and ERD
Mand	- mandatory with sub-typing	Opt	- Optional attributes
*HPKM	– high previous knowledge of method		
* LPKM	– low previous knowledge of method		

The important hypotheses for this study are contained in the columns labeled H1.2 and P2.1. These columns are important as they represent direct comparisons between alternative analysis grammars. Of the predictions, the top three predictions (Comprehension, problem solving, and Cloze test) are the most important as these are measures of learning performance.

Table 3 will serve as a touchstone for the remainder of the thesis, as we will refer to it on several occasions. Having outlined the hypotheses and predictions in this chapter, we now turn to a description of the methods used in the studies in Chapter 5.

Chapter 5: Method

5.1 Introduction

Two studies will be described in this chapter. One study uses the instrument adapted from Mayer to compare the level of understanding and ease of use associated with text representations, data flow diagrams (DFD) and entity relationship diagrams (ERD), and object oriented diagrams (OOA). This is the intergrammar study. A process tracing sub-study was also undertaken in the intergrammar study in order to gain an improved insight into the process of participants used to developing understanding. The methods used in this sub-study will be discussed in a separate section. The second study, the intragrammar study, revisits the research developed by Bodart & Weber (1996) to compare alternative entity relationship diagram grammars. Mayer's methods are also adopted in the second study and will be described below.

5.2 Study 1: Intergrammar Comparisons

The purpose of this study is to compare the comprehension, problem solving, and text reconstruction performance of participants who are presented with one of three representations of the same business domain. The three presentation methods used were text (TXT), data flow diagrams (DFD), and object oriented analysis (OOA). The sections below outline the empirical procedures used in the study.

5.2.1 Method

Participants

One hundred and six students from the University of British Columbia took part in the study. The group was made up of 95 undergraduate students in their third or fourth year and 11 MBA students. The MBA students were distributed as evenly as possible across the three treatment groups (4 in two groups and 3 in a third group). Females accounted for 62 of the participants, males the remaining 44. Sixty-eight of the participants had taken a course on system analysis and had been exposed to the methods used in the study. All participants had previously completed an in depth database assignment using Microsoft Access. Participants were asked to complete two cases and received $15 for completing the study. The average time to complete the study was 81 minutes. The minimum time was 35 minutes the maximum time was 141 minutes.

The study was conducted in a quiet room with 8 computers available for the testing. Groups of from 1- 8 people took the test simultaneously in this room. In the event that only a single person was scheduled for the study, the study was conducted under similar conditions in the researcher's office. Each participant was provided with a computer and instructed to complete the test independently. No talking between participants was allowed. The researcher monitored the room during the study and answered questions regarding the operation of the program. No help was given to any of the participants by the researcher in regards to questions on the test.

Design:

The participants were randomly assigned to one of three treatment groups. One treatment group received a text description (TXT) of the two cases, another group received both a data flow diagram (DFD) and a related entity relationship diagram (ERD) (two separate pages) for each case. A third group received an object-oriented diagram (OOA) (one page) for each case. All participants completed two cases as tasks: Organizing an IFIP Conference (IFIP) and Organizing and Entertainment Event (Event). The first case was created from the IFIP 8.1 Working Group example case. The second case was based on a work description provided by a sports and entertainment franchise. The text descriptions were used as the basis for creating the OOA diagram and the DFD and ER diagrams for both cases in the study. Two cases from two separate sources were used to provide improved external validity of the comparison between grammars. Appendix A provides the text description and all related diagrams for both cases in this study.

Two cases were used in the study in an effort to improve the external validity associated with the findings. This is best illustrated by an example. It is difficult to defend a significant effect arising from a single case since the findings may be case dependent. Finding a similar effect in two cases, reduces the impact of case sensitivity on the validity of the results. The use of two cases also helps to reduce the issues of question bias that may affect findings in a single case. The use of multiple cases is suggested by Bodart & Weber (1996) and will be followed in all of the experiments in this research.

Materials:

The paper materials consisted of an information sheet and consent form, a page of directions related to the test and ten 8.5 x 11 problem solving sheets (5 for each case). The paper materials were handed to participants in an envelope. Each envelope also contained descriptions (either text or diagrams) of two cases. Appendix A has an example for each of the case descriptions. If a participant received a diagram, one or two pages[8] of explanative material outlining the symbols used in the diagram along with a small case example were given to each participant. No training in the method was provided to any participants.

Each participant was also provided with a 3.5" disk that was used to collect information from a testing program written in Microsoft Access version 2.0. No knowledge of Access was required to run the program. Screen captures of the various forms used to run the program are provided in Appendix B. A testing program collected answers as well as elapsed time for all parts of the test. Participants were told that the test was timed. Each participant began the test by registering his or her name and choosing the appropriate test version. The test version was displayed on the envelope containing the paper materials.

Presentation Methods

The text version (TXT) of the "Organizing an IFIP Conference" was provided by the IFIP 8.1 working group example. The "Organizing an Entertainment Event" case was created for this study, and has not been used in other studies. The data flow diagram (DFD) for each case was created using the grammar specified by Gane & Sarson (1979). A related entity relationship diagram (ERD) was created for each data flow diagram using the grammar outlined in Chen (1976). The object-oriented diagrams (OOA) were created using a grammar specified by Wand and Woo (1993) that is similar to the object-oriented method proposed by Coad and Yourdon (1992). These methods have been described earlier in Chapter 4.

The diagrams were created using an iterative process. The text descriptions for both cases were created first. Next, these descriptions were provided to three experts in system analysis methods who were comfortable with the grammars used for the DFD, ERD, and OOA model. The diagrams were collected from these three individuals, and after consolidating the representations, a new diagram was created. This new diagram was then shown to the modeling experts and some further refinements to the diagrams were made. Before the

[8] The text description required no legend explaining the grammar used. Participants received a one page legend, including example, for the symbols used in OOA . Participants using DFD and ER received two legends , including an example, for both the DFD and ERD. See Chapter 4 Diagram 5,6,7, and 8 for legends.

diagrams were used in the study, all experts agreed that the diagrams provided a satisfactory description of the original case[9].

Procedure:

The Microsoft Access testing program automatically led each participant through all portions of the test. The study began with the participant registering for the test by typing their name and selecting the version of the test they were using. The version was written on the outside of the envelope. Next, a pretest asked participants to respond to questions regarding their experience with modeling methods and knowledge of the domains used in the two cases (entertainment events and an academic conference).

To assess each participant's knowledge regarding the presentation methods, participants were asked about the three presentation grammars separately. Participants were asked if they had used the method, how long they had been familiar with the method, and to rate their competence, confidence, and familiarity with each method on a 7-point scale from very low (1) to very high (7). To assess participant's knowledge of the domain, participants were asked to rate their level of familiarity with *"Organizing an academic conference"* or "*Organizing a sporting event*" on a 7-point scale from very low (1) to very high (7). Participants were also asked to "*place a check in the box next to the things you have done*" for 10 activities (5 activities related to each case). Each check box counted as one point. The pretest questions are provided in Appendix B.

After the pretest, the participants were moved though three different tasks for each case. Each participant went through the entire set of tasks for the first case, and then repeated the same tasks for the second case. The order of the cases was randomized within groups based on the version of the test the participant was provided with. The first task was a comprehension test. In this section participants were asked to freely interact with the diagram or text they were given and to answer questions about the case. Questions were the same for all treatment groups. Participants had to answer either "Yes", "No", or "Uncertain" to the comprehension questions. There were 12 comprehension questions in this task. The questions are provided below in Table 6 and again in Appendix C. Questions were designed to move participants through the entire diagram as well as to determine participant's comprehension of the diagram. Participants were not restricted to answering the comprehension questions in a particular order. Participants were also able to scroll back through the questions and to change

[9] A special thanks to William Tan and Darrell Jung for their help in developing the diagrams.

their answers. Average time to complete the comprehension task was approximately 8.5 minutes per case.

An important issue regarding the questions is the potential for biasing the result of the research by choosing questions that "lean" towards a particular method. To reduce the potential for this effect, the questions used in the intergrammar comparison had been used in a previous pilot study that included 91 participants. Some of the questions on the initial instrument were changed based on feedback from participants. The researcher developed all of the questions for the pilot study. Each question was designed so that the question could be answered using any of the presentation methods. An effort was made to move the participant through the entire description, so that participants were likely to be familiar with the entire description before moving into the problem solving questions. In the pilot study, participants were given either a text description, a data flow diagram (with **no** Entity Relationship Diagram), or an Object-Oriented Diagram. In regards to the comprehension questions, no significant difference was found between the comprehension scores in any of the three participants groups (TXT, DFD, or OOA). This finding suggest that there was no systematic bias in regards to the total score attained in the comprehension test. It also indicates that the entity relationship diagram was not essential to score adequately on the comprehension test. The questions are listed below in Table 4.

Following the comprehension test, the participants were told to put away the diagram or text they used in the first section. **The diagram or text description was not available, therefore, to participants for the problem solving task**. Participants were forewarned that they would not be able to view the description after completing the comprehension test. The procedure described in Mayer (1989, 1990, and 1992) also did not permit participants to view the models during problem solving. Mayer also provided less than 8 minutes for participants to view the model. On average, participants took over 8 minutes to complete the comprehension task, so the viewing time is consistent with Mayer's technique.

Removing the diagram meant that participants had to rely solely on their internally developed models (cognitive models) of the system to answer the problem solving questions. Removing the diagram removed the potential confounding effect that the diagram or text description might have on the problem solving process. If the diagram had not been removed, it would have been impossible to identify which solutions were developed from the cognitive model, and which solutions had been generated by further scanning the diagram or text description.

Table 4-A: Comprehension Questions for Study 1

Ques	IFIP Case	Event Case
1	Is the organizing committee responsible for preparing the final list of participants?	Is the operations department responsible for deciding when events are to be held at the complex?
2	Does an author know if their paper has been accepted by the program committee?	Does a promoter know that their event application has been accepted and placed into the program of events?
3	Is the organizing committee responsible for developing the conference itinerary?	Is the human resource department responsible for creating the deployment sheets?
4	Does a participant have to be an author to be placed on the final list of participants, or are there other ways to participate?	Can an employee be scheduled on days they are not available?
5	Is the program committee responsible for maintaining a list of accepted authors ?	Is the promoter the person who determines the number of employees who should work an event?
6	Does the program committee offer invitations to participants?	Is the operations department responsible for creating the employee schedule?
7	Do the Program Committee and Organizing Committee have to interact with each other?	Do the marketing department and the operations department share similar information?
8	Does the program committee keep track of all the papers that are submitted to the conference?	Is the marketing department responsible for keeping track of all of the events that will be held in the upcoming months?
9	Are all of the papers that are submitted to the program committee accepted by the program committee?	Are all of the seating plans provided by the promoter accepted by the operations department?
10	Can an author also be a part of the organizing committee?	Do all employees work every event?
11	Does the program committee have access to the final list of participants?	Is the operations department responsible for creating a seating plan for every event?
12	Is a paper the first thing that an author sends to the program committee?	Is the seating plan directly used in the development of the employee schedule?

Since the diagram was not used, what was measured in the problem solving task was the internal (cognitive) model that participant's developed from viewing and answering comprehension questions about the diagram or text. Since the objective of the problem solving task was to challenge participants to use their understanding of the system to answer questions not directly answerable from just the symbols on the diagram, removing the diagram provided the measure of understanding this researcher was looking for.

After removing the diagram or text description, the participants were asked to answer five problem solving questions for the case. Problem solving questions were of the form "*An employee was available to work on a Saturday but was not scheduled to work. What could have happened for this to occur?*" Participants were instructed to provide as many possible solutions to the problem that they could think of. The participants wrote their answers on the sheets provided to them in the original envelope.

The researcher developed all of the problem solving questions. In developing the questions, the researcher utilized previous problem solving questions reported in Mayer (1989) and Mayer & Gallini (1992). The development of these questions was by no means systematic. In a small case, the researcher found it difficult to create five questions. Since the questions were developed by the researcher, the issue of bias must be addressed. Care was again taken to make sure that none of the questions could be more easily answered by a particular presentation method. This is difficult, however, to establish. It should be recognized that the ability to bias the problem solving questions is less than in the comprehension tests, as the point of the problem solving test is NOT to ask questions about the description, but instead to ask questions of a conceptual nature that relate to the description. The pilot test results are again useful in addressing this question. None of the 91 participants in the pilot study (or the 227 participants in the two main studies) ever mentioned that a problem solving question "could not be answered". The list of problem solving questions for this case is provided in Table 7 below and also in Appendix D.

Both the comprehension and problem solving questions used in the study were pre-tested in a pilot study involving using 91 students. The questions were refined during the pilot study, and no significant changes were made to the questions after the completion of the pilot study. A discussion of the validation preceding the implementation of the empirical instrument described in this section is left to Chapter 7 where the validity of the empirical instrument is discussed in greater detail.

Table 4-B: Problem Solving Questions for Study 1

Ques	IFIP Case	Event Case
1	Is it possible for the same individual to receive two invitations to an IFIP conference? If you answer yes, indicate how this duplication might occur. (Provide as many suggestions as you can think of). If you answered no, indicate why it cannot happen.	Suppose that an employee was available to work on a Saturday but they were not scheduled to work. How could this have occurred?
2	Suppose that a person was invited to submit a paper but has not yet received an invitation. What could have happened?	Suppose that it is a very busy two weeks for the sport complex with events scheduled for every day of the week. What problems might be encountered due this busy schedule?
3	Neither the program committee nor the organizing committee is directly responsible for maintaining the mailing list of potential authors and participants. What organization problems do you think could arise from this situation.?	Suppose that a promoter underestimated the number of people who wanted to attend a event. The promoter now wants to increase the number of seats available for an event. Given that there are some seats that could be opened, what problems would be created by opening more seats?
4	Suppose that a very highly respected author indicated that she would like to present a paper in the conference but she has given the program committee only two weeks notice. What problems would be faced by the committees in accepting the author's wish?	Some events take a long time and a large number of people to setup and take down (this is referred to as conversion time). What organizational problems can be caused by events with a long conversion time?
5	Suppose that instead of inviting specific authors to submit papers and inviting only selected participants to come to the conference, the two committees agreed to open participation to any individuals and all interested authors. What problems would arise if the current system was used to organize the conference?	Assume that the marketing department and human resources department act independently of each other. What problems might occur if the se two departments do not share their information quickly or work closely with the other department?

Upon completing the problem solving task the participants were asked to complete the "fill-in-the-blank" or Cloze test. This test asked participants to fill in blanks left in the text used to create the original models. A total of 45 words were replaced with blanks from a total of 353 words in the description. The words to be eliminated were chosen by the researcher on the basis of their importance in the description. Once a word had been eliminated and replaced with a blank - for example the word "deployment" - all of the instances of that word were eliminated from the text description. Participants were instructed that each blank was associated with one word and one word only. They were then asked to fill in as many of the blanks as they possibly could.

The Cloze test was treated in a similar way to the problem solving task, in that **participants were told not to use the model or text originally provided and to fill in the Cloze test answers**. Two reasons are provided for removing the diagram or text description. First, in the case of the text description, the Cloze test would become simply a "copying" exercise rather than a test of knowledge. With the text description in front of them, the Cloze test would be meaningless. Second, as noted earlier the Cloze test was designed to test semantic information rather than short term memory. Providing the text description or diagram would have made it difficult to separate the semantic knowledge from the information simply provided on the diagram. The removal of the description provided a clear test of the understanding each participant had developed. Removing the text description or diagram, therefore, improved the quality of the measure of understanding that was developed by the participant viewing the model. An example of a portion of the Cloze test is provided in Figure 9 below. Appendix E contains the full version of both Cloze tests used in this study.

After completing the first "fill-in-the-blank" test, each participant was walked through the same set of three tasks with a second case. A post-test followed the completion of the second case. The post-test was used to capture the perceived "ease of use" associated with a modeling grammar. The perceived "ease of use" measurement was based on the ease of use instrument developed in Moore and Benbasat (1992). The four questions related to ease of use are provided below in Table 8 and also included in Appendix F.

Figure 9: Sample Cloze test for Study 1

Organizing an Entertainment Event

A company that owns a large sports and entertainment complex holds a variety of _______________ in their building including professional sports, concerts, and conventions. The _______________ of events is planned several months in advance. While no more than one event can take place at any given time, it is possible for more than one _______________ to occur on the same day.

To book an event, a _______________ contacts the _______________ department for the complex and explains the type of event being promoted, the expected _______________, and a requested date and time for the event. After reviewing the _______________ application, the _______________ department either accepts the _______________ and places the event in the _______________ of events, or rejects the application. Once an event _______________ is accepted, the _______________ is responsible for sending a _______________ plan for the event to the _______________ department of the complex.

The operations department reviews all _______________ plans and decides on the appropriate number of security personnel, _______________, and customer service representatives for the _______________. These estimates for _______________ requirements are _______________ with the _______________ before they are finalized. Once the _______________ requirements are finalized, a _______________ sheet is created by the operations department. The _______________ sheets lists the start time, end time, location, and _______________ requirements for every _______________ that will be required for the _______________. The completed _______________ sheets are then passed to the Human resource department who is responsible for _______________ staff to work the event.

Table 5: Ease of Use Post-test Questions for Study 1

Ease of Use Questions	**Description** (replace *********** with either text description, OOA, or combination of DFD and ERD)
1	I believe that it was easy for me to understand what the *************** was trying to model.
2	Learning how to read the ************* was easy for me.
3	Overall, I believe that ************ is easy to use.
4	Using the *********** method was often frustrating.

Scoring

Three dependent measures of the "product" were collected: comprehension score, problem solving responses, and Cloze score. Measures of the learning process such as perceived ease of use and completion times for the comprehension, problem solving, and Cloze tasks were also collected. The comprehension score provided the number of correct responses from a total of 12 questions. Correct answers were determined using the text descriptions. The comprehension scores ranged between 4 and 11.

To create a score for the problem solving, Mayer's (1989) procedure in creating a set of "acceptable" responses for each of the five problem solving questions was followed. For example, acceptable answers for the "***why an employee was not scheduled to work on a Saturday***" question noted above included: "*the employee did not have required skills*"; "*staffing requirements did not need that many people"; "employee has already worked too many hours*"; "*there was no event on a Saturday*"; "*employee data was not up to date*"; and "*employee did not hand in availability card.*" This list of acceptable answers for all problem solving questions are provided in Appendix D. The list of acceptable answers was determined between the two coders, with each coder developing suggestions. The problem solving score is created by summing of the number of "acceptable solutions" generated by a participant across the five problem solving questions. Two coders created independent ratings of the scores[10]. The Pearson correlation coefficient for the ratings created by the two scorers was over 0.90 for all cases. A more detailed analysis is provided in Chapter 6. The number of acceptable responses ranged between 7 and 35.

The Cloze score was created by summing the number of exact words or synonyms that were correctly filled-in by participants from a total of 45 blanks. The list of acceptable synonyms is provided in Appendix E. Cloze test scores ranged between 10 and 43. Finally, the perceived ease of use score represent the summation of responses to four questions rated on a scale from 1 (strongly disagree) to 7 (strongly agree). The summation of the four items in the perceived ease of use ranged between 13 and 26.

5.3 Process Tracing for Study 1

Initial results from the empirical study described above provided preliminary evidence of significant differences in the understanding developed by participants across the presentation

10 Thank you to Nandakumar Sivakumar who independently rated the scores

methods. The empirical method described in the previous section outlined a procedure for measuring differences between presentation methods. These procedures, however, are not able to address the question of **why** these differences occurred. To address why understanding is affected by different presentation methods, it is necessary to trace the process of how participants develop their understanding. The measurement of these cognitive activities underlying understanding requires methods that shed light on the process of participants' use in developing an understanding of the domain presented to them. The empirical methods used to observe cognitive processes are often referred to as **process tracing methods.** These methods provide, as Todd & Benbasat (1987, p. 495) suggest, "*...some access to activities that occur between the onset of a stimulus and the eventual response or choice.*"

Why use Verbal Protocols?

A variety of process tracing methods have been developed including tracking eye movement, computer logs, written protocols and verbal protocols. This study will make use of verbal protocols over the other techniques. Verbal protocols have been chosen for three reasons. First, they provide a method for assessing the understanding process concurrently, that is, while participants' are actively developing understanding. As Ericsson and Simon (1993, p. xii) note:

> *"There is a dramatic increase in the amount of behavior that can be observed when a subject is performing a task while thinking aloud compared to the same subject working under silent conditions. A brief instruction to think aloud usually suffices to bring about this major change in observable behavior. In light of the fact that subjects do not need to practice before being able to "think aloud" one infers that this verbal reporting is consistent with the structure of their normal cognitive processes and general skills for verbalizing needed information. "*

This type of observation method provides the capability of a more direct observation of the learning process.

A second reason for adopting verbal protocols is the precedence already created for the use of concurrent verbal protocols in systems analysis research, (Jarvenpaa & Machesky, 1989; Vessey and Conger, 1994; Siau, 1997) and in the area of understanding (Ericsson & Simon, 1993). Researchers have used the method to successfully differentiate between presentation methods. This suggests that the verbal protocol observations are capable of producing data that could be used to address questions as to why there are differences in understanding across the three presentation groups.

5.3.1 Method

Subjects and Design

The effort involved in analyzing data from verbal protocols generally precludes large sample sizes. Participants included 12 students from the University of British Columbia. The participants included 7 females and 5 males. All students were at the Masters level and had taken at least one management information systems course. Of the 12 participants, 6 had taken a systems analysis class and had developed knowledge of the modeling techniques used in the study. The remaining 6 participants had very little or no knowledge of the modeling techniques. Masters level students were chosen for the sample because they participated previously in the larger experiment, for their a priori ability to articulate their thoughts (to think aloud), and the fact that the Masters level students were considered representative of many users in information systems domain

Three presentation methods were used in the test with four participants in each group. Two of the four participants in each group had prior knowledge of analysis, and two did not. Each participant was paid $20 for his or her time. Only one case was used in the study (the Event Case). The Event case was chosen because, in preliminary analysis, the case provided the largest difference in scoring between the methods. Since the differences were more apparent in the scoring, it was thought that the Event case would provide the best opportunity for observing differences in the interpretation process. The average time to complete the process tracing task including comprehension, problem solving, and Cloze task was approximately 49 minutes.

Materials

The materials provided to the process tracing participants were the same as those described in the previous section. The paper materials consisted of an information sheet and consent form along with a page of directions related to the test. The legend and sample case was also provided to the participants.

Each participant was also given a 3.5" disk that was used to collect information from a testing program written in Microsoft Access version 2.0. This testing program was the same used in the earlier section. No knowledge of Access was required to run the program. Screen captures of the various forms used to run the program are provided in Appendix H. The testing program collected answers as well as elapsed time for all parts of the test. Participants were told that the test was timed.

Each participant began the test by registering his or her name and choosing the appropriate test version. The text version was displayed on the envelope containing the paper materials. The sessions were held in a secure, quiet office. The researcher was present for the entire session for all of the participants, and participants were tested one at a time. Each participant was provided with a laptop computer that held the testing program. The sessions were video and audio taped. The camera was placed to the side and slightly behind the participant so the camera was not in the participant's direct vision. The camera was angled to show a slightly "over-the-shoulder" view of both the participant and the computer screen. Microphones were placed in front and to the side of the participant so that the microphone was less intrusive.

Each session began with a short description of what the subject could expect, and the "talk aloud" protocol. When participants were briefed and comfortable, they were asked to complete a pre-test. After the pretest, the participants were guided through the comprehension, problem solving, and Cloze test, in a similar manner as described above in Section 5.1.1. The one large difference was that participants were asked to "think aloud" as they answered the questions. The "think aloud" procedures are described in the next section.

Procedure

The study began with the pre-test questionnaire that was completed on the computer. The pre-test questionnaire was the same as described in Section 5.1.1. Following the completion of this questionnaire, the participant was handed a description of a domain using one of the modeling techniques (text, DFD/ERD or OOA). Each participant was then instructed that they should "think aloud" as they answered the questions they were presented with.

In eliciting the protocols, the ***active probing*** method as categorized by Todd & Benbasat (1987) was adopted. In this method, the participant is asked to "think aloud" while they are engaged in answering a question. No specific instructions on how to proceed are given to the participant by the researcher. The researcher will prompt a participant with statements such as "*What are you thinking?*" if an extended silence (more than approximately 10 seconds) is observed. Also, in points where the researcher sees an important issue developing, the researcher can ask questions such as "*Why did you choose that?* or "*Why is* *that important*?" The active probing method can be intrusive on the learning process if the probing questions are used often, however this method does produce more protocol information in the areas where the research is focused

After receiving the instructions to "think aloud" the participants were asked to complete the comprehension questions. These questions provided the stimulus necessary for the collection of verbal protocols. Several individuals were initially uncomfortable with "thinking aloud", however, all of the participants were capable of providing useful protocol data and felt comfortable after answering only a few questions in the comprehension test.

The comprehension test was followed by the same problem solving tasks and the Cloze test as describe in Section 5.1.1. Participants were encouraged to "think aloud" for both of these tests. As was the case in the initial study, the diagram or text description was taken away from the participant before they started answering the problem solving and Cloze tests. The participants, therefore, had to rely solely on their cognitive models to complete these tasks. In the interest of keeping the protocol data as "clean" as possible, the participants were told that they did not have to write anything down for the problem solving task. The writing, it is assumed, would have slowed down or altered the cognitive process. In a similar fashion, participants were also asked to simply verbalize answers to the Cloze test rather than typing the answers into the computer. After completing the Cloze test, the participants were asked to complete a post-test questionnaire that contains the ease of use scale. No verbal protocols were collected for this section.

The final task in the protocol analysis was for each participant to create their own diagram describing the domain they had just learned about. The participants were read the following paragraph, and then asked to draw:

> **Assume that a person will be coming into this room and that your job is to explain to this person the case that you have just learned about. You can assume the new person has no knowledge of the case and no knowledge of any formal system analysis methods or technique. In preparation for this explanation, I would like you to create a diagram that will help you explain the system to the new person.**

Participants were told they were free to draw any diagram that they wanted. They were not encouraged to redraw the diagram they had been given. The results of the drawings are presented in Appendix G. The completion of the diagram, indicated the completion of the participation in the study.

Coding Schemes

Collecting verbal protocol data can provide a researcher with potentially valuable qualitative information. Once collected, the protocol data must be coded in order to extract this value. Opinions differ in the approach that a researcher should take in extracting information from protocols. For example, Todd & Benbasat (1987, pp. 499) suggest that:

> *"... protocol coding schemes should be created, at least in part, **a priori**. A priori determination ensures that the "findings" of the study are not data-driven. Strict independence is maintained between hypothesis formation and data analysis."*

This view of *a priori coding* can be contrasted with other forms of qualitative data analysis such as the "grounded theory" approach suggested by Strauss & Corbin (1990, p. 95) who state:

> *"Each of us brings to the analysis of data our biases, assumptions, patterns of thinking, and knowledge gained from experience and reading. These can block our seeing what is significant in the data, or prevent us from moving from descriptive to theoretical levels of analysis. "*

The two approaches to qualitative data analysis noted above suggest two very different methods for coding the verbal protocol data. An important questions to consider, therefore, in the analysis of protocol information is "What approach will be taken to analyze qualitative data?"

The choice in approach is dependent on the questions being considered in the research. The focus on empirical comparisons between methods suggests an admitted quantitative bias on the part of the researcher. This bias suggested that more objective measures would be preferred over less objective measures. The bias also implied that the analysis of protocol data, in this thesis, would be used to support and qualify the findings in the larger quantitative study described earlier in section 5.1. The protocol data in this thesis, therefore, is not being used to develop theory. Instead, the data is being used to explore differences in cognitive processes. These statements suggest that this study will lean more towards the a priori coding schemes suggested by Todd & Benbasat (1987) than the grounded theory approach suggested by Strauss & Corbin (1990). The approach used in this thesis relegates the protocol data to a supporting role, where the protocols can be used to triangulate quantitative findings. The supporting role for protocol represents a pragmatic choice to focus resources on the quantitative studies which more closely reflect available research skills and experience than a grounded theory approach.

Coding

Since the protocol information was being used primarily for triangulation and anecdotal evidence, the analysis of the protocol information was left until after the quantitative study was completed. The results in the quantitative study, therefore, provided a guide for the analysis of protocol data. For this reason, much of the discussion of coding is left until a discussion of the results in Chapter 7. Some issues describing the type of coding is discussed below.

A part of the coding scheme was based on the cognitive activities required to develop understanding. These activities include selection, organization, integration, and encoding of information embedded in the text or diagram. These activities have been noted previously in Chapter 3. It was immediately obvious from the outset of the study that the protocol data would not be able to provide useful information that was able to separate and identify these cognitive activities. From this recognition, and the interest of developing objective measures from the qualitative data, the idea of measuring pauses and elapsed times to answers was created.

All protocol data for the problem solving task were split into 5 second intervals and coded for pauses and acceptable solutions. The five-second interval was chosen primarily as a matter of convenience and worked well to split the protocol data into even units of time. The 5 second interval was the first dimension coded. Marks were placed directly on the paper transcription of the protocol at five second intervals. These marks provided useful guides to the time spent on answering questions.

Pauses were the next dimension coded. A mark was placed in the transcript at the time of the pause to indicate the person had stumbled or stopped at that point. A **pause** was defined as a silence that lasted more than second, or a repeated or stalling word (such as umm, well, ahh, hmm, etc.). The researcher did the coding for all of 12 of the participants. A second independent coder was also used to validate the coding procedure[11].

On a third pass through the data, the acceptable answers were highlighted in the text and the number of acceptable answers was counted for all questions for each participant in the

[11] My thanks to Stacey Rain, MBA student at Simon Fraser University for the difficult and timely coding of protocols.

test. These three dimensions (time, pauses, and acceptable answers) provided an *a priori* coding scheme that was able to deliver relatively objective results from the protocol data. Several other measures were also developed after some quantitative data analysis. This data and the subsequent data that flowed from the protocol analysis is provided in Chapter 7.

5.3 Summary of Intergrammar Study

The previous sections in this chapter have outlined the empirical methods that will be used to address the first two hypotheses in this thesis. An experiment designed to compare text representations, a combined DFD/ERD method, along with object oriented analysis provides the main body of evidence that will be used to test the hypotheses. A follow-up protocol analysis is used to provide triangulating data has also been described. Together these two empirical procedures should provide data for an "interesting" comparison between the three presentation methods.

5.4 Study 2: Mandatory and Optional Properties – An Intragrammar Study

The purpose of this study is to revisit Bodart & Weber (1996) using the same empirical methods described in Study 1 above. The original study hypothesized that in the use of the entity relationship diagram (Bodart & Weber, 1996, pp. 450):

> "*the use of subtypes with mandatory properties in conceptual schema diagrams will communicate the meaning of the real-world domain to users better than the use of optional properties in conceptual schema diagrams.*"

Bodart & Weber's theoretical arguments are based on the Bunge-Wand-Weber (BWW) ontology (as discussed previously) which can be used to suggest that ERD representations using mandatory properties are ontologically superior to ERD representations using optional properties. Unfortunately, the results obtained through their free-recall procedure, while promising, did not provide convincing support for their theory. Their study is revisited here in the hope that the instrument suggested in this thesis provides the necessary empirical sensitivity to answer the theoretical questions posed by Bodart & Weber (1996).

5.4.1 Method

Participants

One hundred and nine students from the University of British Columbia took part in the study. Of these 109 participants, 99 were undergraduate students in their third or fourth year and 10 of them were Masters students. The participants included 58 females and 51 males. All of the participants had taken a course in management information systems, and 53 of the participants had taken (or were taking) a course on system analysis and had some knowledge of the entity relationship model. None of the participants were familiar with the optional and mandatory constructs as described in Batini (1992). All participants had previously completed an assignment using Microsoft Access. Participants were asked to complete two cases and received $15 for completing the study. The average time to complete the study was 88 minutes.

Design

The participants were randomly assigned to one of three treatment groups. One treatment group received a text description of the two cases, one group received a model developed using optional properties for each case, and a final group received a model using subtypes with mandatory properties for each case. The text description served as a "control" for the graphical cases

Two entity relationship modeling grammars were used in the study. The alternative grammars are outlined in Batini (1992) and illustrated in Diagram 8 in Chapter 4. The entity relationship model (ERM) was a good choice for the study for three reasons: the popularity of ERM with system analysis practitioners; the availability of grammars using both optional and subtype with mandatory properties; and the familiarity with ERM among almost half of the study participants.

All participants completed two cases: Voyager Bus and Far East Repair. The first case, Voyager Bus, was created and used by Bodart & Weber (1996) and is repeated here. The second case is derived from the Far East case provided in Batra, Hoffer, & Bostrom (1990). The same three person group of "experts" was used to develop and validate the ERD for the Far East Repair case. Since previous empirical comparisons have noted the sensitivity of results to the case used in the study, two cases from two separate sources were used to provide improved external validity of the comparison between grammars. Appendix A provides more information regarding these cases.

Materials

The materials used were the same as those described in Study 1. The paper materials consisted of an information sheet and consent form, a page of directions related to the test, and ten 8.5 x 11 problem solving sheets (5 for each case). The paper materials were handed to participants in an envelope. Each envelope also contained descriptions (either text or diagrams) of two cases. Appendix D has an example for each of the case descriptions. If a participant received a diagram, a one page description of the grammar (similar to that shown in Diagram 8 in Chapter 4) was also provided.

Each participant was also given a 3.5" disk that held a testing program written in Microsoft Access version 2.0. The program collected answers as well as elapsed time for all parts of the test. Each participant registered by inputting their name and the test version displayed on the envelope. No knowledge of Access was required to run the program. The sessions were run in a quiet room with between 1-8 persons. Each participant was provided with a computer. Participants worked independently and the researcher supervised all of the sessions. Screen captures of the program are provided in Appendix B.

Procedure

The program automatically led each participant through all portions of the test. The study began with a pretest that asked participants to respond to questions regarding their experience with entity relationship modeling and the subject domain of the two cases (a bus company and an engine repair company). The pretest is provided Appendix B. The time taken to complete each section was recorded and participants were told the tests were timed.

After the pretest the participants were moved though three different tasks for each case. Each participant went through the entire set of tasks for the first case, and then repeated the tasks for the second case. The order of the cases was randomized within groups. The first task was a comprehension test. In this section participants were asked to freely interact with the model or text they were given and to answer questions about the case. Participants had to answer either "Yes", "No", or "Uncertain". Example questions included "*Are all buses that are available assigned to a daily route segment?*", or "*Can a trip be made up of more than one* *route segment?*" There were 12 questions in this section. Appendix C contains the entire list of questions used in both cases.

As noted earlier, there is potential for bias in these questions. Ten of the questions used for the voyager case were developed by Bodart & Weber (1996) and used directly. All of the questions for the Far East case were developed by the researcher. As usual, care was taken to make sure that questions could be answered using any of the presentation methods. Since the comprehension scores showed little significant difference between the treatment groups, the researchers shows no signs of systematic bias in the questions. The questions are listed below in Table 6 and again in Appendix C.

Following the comprehension test, the participants were told to put away the model or text they used in the first section. They were then asked to answer five problem solving questions for the case. Questions were of the form "*A bus driver for the Voyager bus company has a problem. All seats on the bus have been taken, yet there is a passenger waiting to board the bus. What could have happened to cause this problem? Write down as many possible answers as you can think of.*" The participants answered these on the sheets provided to them in the original envelope.

The researcher developed all of the problem solving questions for the two cases used in the intragrammar study. In developing the questions, the researcher utilized previous problem solving questions reported in Mayer (1989) and Mayer & Gallini (1992). Since the questions were developed by the researcher, the issue of bias must be addressed. Care was again taken to make sure that none of the questions could be more easily answered by a particular presentation method.

Table 6: Comprehension Questions for Intragrammar

Question: Voyager Bus		Answer
1.	Can a trip be made up of more than one route segment?	Y
2.	Does a person have to make a reservation to go on a trip?	N
3.	Can a daily trip be assigned to more than one bus?	Y
4.	Does Voyager Bus Inc. collect the same set of information for all of the passengers?	N
5.	Can the same daily route segment be associated with two different trip numbers?	Y
6.	Can Voyager Bus Inc. record a maintenance problem that has not yet been experienced by any of their buses?	U
7.	Is the daily route segment modeled as an entity?	Y
8.	Can a bus driver be assigned to more than one daily trip?	Y
9.	Are all buses that are available for use assigned to a daily route segment?	U
10.	Is model an attribute of bus?	Y
11.	Is the average cost of repair recorded for all buses in maintenance?	Y
12.	Can the end town assigned to a route segment be different from the end town associated with a trip that uses the route segment?	Y
Question: Far East Repair		**Answer**
1.	Do all repairs require parts?	N
2.	Can a repair be worked on by more than one mechanic?	Y
3.	Are all repairs assigned to at least one mechanic	Y
4.	Are there parts stored in the warehouse that are not used for repairs?	U
5.	Does Far Eastern collect different information for different machine types?	Y
6.	Does Far Eastern differentiate their local customers in any way?	Y
7.	Can a mechanic who does not have a special skill be assigned to more than one repair?	Y
8.	Do all the mechanics related to the same repair, pool their hours to create a single entry for hours worked?	N
9.	Can a piece of equipment undergo more than one repair?	Y
10.	Can more than one part be listed in a single repair detail?	Y
11.	Is the cylinder volume recorded for all pumps that are repaired?	N
12.	Can a part be supplied by more than one manufacturer?	Y

Upon completing the problem solving task the participants were again told not to use the model or text originally provided and to complete the "fill-in-the-blank" or Cloze test. This test asked participants to fill in blanks left in the text used to create the original models. A total of 45 words were replaced with blanks from a total of 350 words. Appendix E contains an example of this test.

After completing the first "fill-in-the-blank test, each participant was walked through the same set of three tasks with a second case. A post test followed the completion of the second case. The post test was used to capture the perceived "ease of use" associated with a modeling grammar. The perceived ease of use measurement was based on the ease of use instrument developed in Moore and Benbasat (1992). The post test is provided in Appendix F.

Scoring

Four dependent measures were collected: comprehension score, problem solving responses, Cloze score, and perceived ease of use. Completion times for the comprehension, problem solving, and Cloze tasks were also collected. The comprehension score provided the number of correct responses from a total of 12 questions. Correct answers were determined using the text descriptions. The comprehension scores ranged between 3 and 11.

To create a score for the problem solving, we followed Mayer's (1989) procedure in creating a set of "acceptable" responses for each of the five problem solving questions. For example, acceptable answers for the "*why there are no seats for a waiting passenger*" question noted above included: "*passenger without a reservation boarded bus*"; "*central scheduling overbooked bus*"; "*wrong bus assigned to route*", "*some passengers boarded wrong bus*", "*bus has faulty seat*", *and* "*one person may take up more than one seat*". The problem solving score represents the summation of the number of "acceptable solutions" generated by a participant across the five problem solving questions. The number of acceptable responses ranged between 3 and 23.

The Cloze score was created by summing the number of exact words or synonyms that were correctly filled-in by participants from a total of 45 blanks. The Cloze score ranged between 13 and 43. Finally, the perceived ease of use score represents the summation of responses to four questions rated on a scale from 1 (strongly disagree) to 7 (strongly agree). The summation of the four perceived ease of use measures ranged between 13 and 26.

5.5 Summary

This chapter has outlined the empirical methods that were used to address the hypotheses raised in Chapter 4. The intergrammar study adapts Mayer's empirical instrument to compare the level of understanding and ease of use associated with text representations, data flow diagrams, and object oriented analysis diagrams for set of participants with a wide variety of domain and modeling experience. The process tracing methods utilizes verbal protocols to observe participants involved in the interpretation of system representations. The data from these protocols may provide useful comparisons of the cognitive processes associated with different modeling techniques. The intragrammar study revisits a study conducted by Bodart & Weber (1996) that compares different representations using two entity relationship diagram grammars. This replication provides a test for the sensitivity of the instrument suggested in this study.

Now that the methods have been outlined, the results of the tests can be presented. These results are developed in Chapter 7. Preceding the discussion of results in Chapter 7 is the description of the statistical techniques and preliminary statistics in Chapter 6. The developments in Chapter 6 are necessary for the results presented in Chapter 7 can be inferred. Those readers unfamiliar with Multivariate Analysis of Covariance (MANCOVA) and the tests associated with the assumptions underlying MANCOVA are encouraged to move to read appendix K where the MANCOVA procedure is discussed in more detail.

Chapter 6: Techniques and Preliminary Statistics

This chapter provides an analysis of the preliminary statistics associated with the two studies in this thesis. Chapter 7 then follows with a detailed analysis of the main hypotheses. It is my hope that the reader's perseverance will be rewarded in the next two chapters with a deeper understanding of the results of the study. Readers who find their interest waning amid the sea of hypothesis tests are encouraged to seek sanctuary in the summary of this chapter provided in Section 6.5. For those braver souls, our discussion of statistical tests and results begins with Table 7 below. Table 7 provides a listing of all of the variable names used in the study and a short description of the measure associated with the variable.

Table 7: Explanation of Variables Used in the Study

Variable Name	Description
Case Names	
IFIP	Organizing an IFIP Academic Conference
EVENT	Organizing an Entertainment Event
VOYAGER	Organization of a Bus Company
FAREAST	Organization of a Machine Repair Shop
Demographic Variables	
EDU	Education level, 0 – undergraduate, 1 – masters
SEX	Female - 0, Male – 1
PREEVEN	Perceived knowledge of organizing entertainment event (1–low, 7-high)
PREIFIP	Perceived knowledge of organizing conference (1 – low, 7 high)
PREVOYB	Perceived knowledge of bus company (1 – low, 7 high)
PREFARE	Perceived knowledge of machine repair shop (1–low, 7-high)
PREIFIP	Total (max 5) number of activities related to IFIP case
PCHKEVEN	Total (max 5) number of activities related to Event case
PCHKVOYB	Total (max 5) number of activities related to Voyager case
PCHKFARE	Total (max 5) number of activities related to Far East case
PUSEDFD	Previously used DFD for modeling (0 –No, 1 – Yes)
PUSEERD	Previously used ERD for modeling (0 –No, 1 – Yes)
PUSEOOA	Previously used OOA for modeling (0 –No, 1 – Yes)
PCMPDFD	Perceived competence with DFD (1–low, 7-high)
PCMPERD	Perceived competence with ERD (1–low, 7-high)

Table 7 con't Variable Name	Description
PCMPOOA	Perceived competence with OOA (1–low, 7-high)
PCNFDFD	Perceived confidence with DFD (1–low, 7-high)
PCNFERD	Perceived confidence with ERD (1–low, 7-high)
PCNFOOA	Perceived confidence with OOA (1–low, 7-high)
PCFAMDFD	Number of months you have been familiar with DFD method
PFAMERD	Number of months you have been familiar with DFD method
PFAMOOA	Number of months you have been familiar with DFD method
Study Variables (replace ** with: EV – event case, IF – IFIP case, VB – voyager case, FE – Far East case)	
**COMPTOT	Total of correct answers in comprehension task
**PRBTOA	Total of acceptable answers in problem solving task by rater A
**PRBTOB	Total of acceptable answers in problem solving task by rater B
**CLZTOT	Total of acceptable synonyms in Cloze test
**CMPTIM	Total time for comprehension test
**PRBTIM	Total time for problem solving task
**CLZTIM	Total time for Cloze test
Scale Variables	
EASE	Additive scale created from 4 posttest variable
KMETHOD	Additive scale created from pretest "method "questions
KDOMEVEN	Additive scale created from pretest "domain" question for Event case
KDOMIFIP	Additive scale created from pretest "domain" question for IFIP case
KDOMVOYB	Additive scale created from pretest "domain" question for Voyager case
KDOMFARE	Additive scale created from pretest "domain" question for Far East case

6.1 Preliminary Statistics for both Intergrammar and intragrammar Study

Before proceeding with the analysis in Chapter 7, some preliminary statistics need to be considered. These preliminary statistics can be broken into two groups; preliminary statistics common to both studies, and preliminary statistics associated with each study individually. In this section we will consider the preliminary statistics that are common to both studies. These statistics include the analysis of scale variables created for the study, the inter-rater reliability associated with the problem solving questions, and the independence between the two sets of dependent variables (effectiveness and efficiency).

Scale Variables: Knowledge of Domain, Methods, and Ease of Use

We begin our discussion of scaled variables by focusing first on the participant's prior knowledge. Each participant brings with him or her previous knowledge about the domains used in the experiment as well as previous knowledge the presentation methods used to describe the domains. While randomization of participants across groups helps to control for the differences of domain and method knowledge across groups, it is important to understand the role that domain and method knowledge have on the dependent measures. We collect, therefore, several measures of a participant's domain knowledge and method knowledge in the pre-test for both studies. These measures are then combined into a single ***scale variable*** that can be used as a covariate in the main analysis.

Scales in this study are created in the following manner. First, several items that are related to a single construct (such as domain knowledge or ease of use) are collected. The scores for each of the items are then standardized into Z scores. This procedure ensures that scores for each of the items has a mean of zero and a standard deviation of one. After standardizing, the internal consistency of the items chosen for the scale is then assessed using Cronbach's alpha. For a description of this technique see Hare (1992, p. 468). Essentially Cronbach's alpha is a measure between zero and one of the internal consistency of the items used in a scale, based on the average inter-item correlation. The larger the alpha, the more internally consistent the scale. Adjustments are made to items in the scale if improvements in internal consistency are necessary. Once the adjustments are complete, the standardized scores from each participant for each the items in the scale are added together to create the single scale variable that represents the influence from each of the items.

The scale variable for domain knowledge was created from two items. This technique was derived from the technique described in Mayer & Gallini (1990). The first item in the scale was a rating (1-low, 7-high) of the participant's perception of their knowledge of a general activity. Participants were asked *to "Indicate their knowledge of the following activities.*" The activities for the intergrammar comparison were "*Organizing an Entertainment Event*" and "*Organizing an Academic Conference*". For the intragrammar comparison these activities were "*Organizing a Bus Company*" and "*Organizing a Machinery Shop*". The second item of the scale was a count (minimum zero, maximum five) of the number of related activities that a participant had experienced. For example, the main activity "*Organizing an Entertainment Event*" had five related activities. Participants were asked *to "Place an X in the box next to the* *activities you have done*". They were then given five related activities: *"Worked at a large entertainment event*", "*Arranged security for an event", "Delegated work for a group of service employees*", "*Made arrangements for scheduling an event*", *and "Promoted an entertainment event*". The number of boxes that were checked formed the second item.

The two items described above were then standardized and combined into a scale variable. There were four scale variables created; one variable for each main activity. The Cronbach alpha for each of the scale variables is provided below in Table 8. No adjustments were made to the items as there were only two items in the scale

Table 8: Reliability Analysis for Domain Knowledge Scale Variable

Study	Case	Variable Name	Alpha
Intergrammar	IFIP	KNOWIFIP	0.737
Intergrammar	Event	KNOWEVEN	0.746
Intragrammar	Voy Bus	KNOWVOYA	0.438
Intragrammar	Far East	KNOWFARE	0.533

In general, the measure of domain knowledge did not provide a scale variable with satisfactory internal reliability. This suggests that more work needs to be done in developing measures of prior domain knowledge.

The scale variable for system analysis method knowledge was created using nine items. Questions were asked regarding three different analysis methods: data flow diagrams, entity relationship diagrams, and object oriented diagrams. Three questions were asked for each of the three analysis method making a total of nine items for the scale. The first question asked *"How many months have you been familiar with ....*", where the "..." indicates the analysis method used by the participant. The minimum was zero and the maximum was 24 months for this item. The second question asked each participant to "*Estimate the level of competence you have attained with ....*". This item was measured on a 7 point scale with 1- low and 7- high. The final question asked each participant to *"Estimate the level of confidence you have attained with ...*". This item was again measured on a 7 point scale with 1-low and 7- high.

In developing the scale variable for this item there was a choice either to separate the items relating to the three analysis methods and create three scale variables, or to combine the three methods into a single scale variable. I have chosen to combine all three methods in this study. The reason for this is twofold. First, from a pragmatic stance the participants in this study are students. If these students have gained any exposure to analysis methods then they have gained that exposure primarily from the system analysis course taught at the university. This course teaches all three of the methods. So exposure to one of the methods generally predisposes exposure to the other methods. The scale showed a high degree of correlation between items which indicates that this effect may have occurred.

Second, it should be noted that prior knowledge of any analysis method may influence the way that a participant views the domain as it is presented to them. The fact that a student has been trained to view system from a system analysis perspective is an important consideration. The nine-item scale would pick up the influence of prior knowledge of any of the three analysis methods. The nine-item scale is perhaps better described, therefore, as a measure of prior knowledge of analysis methods in general, rather than knowledge of a specific analysis method.

Given the above considerations, a nine-item scale was created from the standardized values. The Cronbach alpha for both the intergrammar and intragrammar study is provided below in Table 9.

Table 9: Reliability Analysis for Scale Variable: Knowledge of Analysis Methods

Study	Type of Knowledge	Variable Name	Alpha
Intergrammar	Knowledge of analysis methods	KNOWMETH	0.872
Intragrammar	Knowledge of analysis methods	KNOWMETH	0.861

The scale variable "knowledge of analysis methods" shows high internal reliability in both studies. Further, the scale is robust in that items deleted from the scale would only serve to decrease the resulting alpha.

The final scale variable to be discussed is the measure for the perceived ease of use of the analysis method used in the study. These items were collected AFTER the test was completed and was the last step before the participant finished the study. Unlike the previous scale variables, the items for this scale were taken from the four-item "short form" version of an 8 item "ease of use" scale previously developed by Moore & Benbasat (1991). These four items were originally designed to measure the ease of use as it related to the adoption of new technology. The items had to be adapted, therefore, to the use of system analysis methods. The four items used in this study were:

1. *"It was easy for me to understand what the ... was trying to model."*
2. *"Overall, I believe that the ... was easy to use".*
3. *"Learning how to read a ... was easy for me".*
4. *"Using the ... was often frustrating."*

Note that in the items above, the "..." would have been replaced with the analysis method that the participant used in the study. The participants were asked to indicate how strongly they disagreed or agreed with the statement (1 – strongly disagree, 7 strongly agree). The fourth item was recoded as it was in a direction opposite to the other three.

The scores for the four items were standardized and then the internal validity of the scale was assessed. The initial assessment showed that the scale could be dramatically improved by dropping the fourth item in the list above. This was likely due to the fact that the fourth item was the last item on the test and that participants may not have been paying particular attention to the change in wording. After eliminating the fourth item, the internal validity of the three-item scale was assessed. Again, the analysis suggested that the validity could be significantly improved by eliminating the third item from the scale. The third item asked participants how easy the analysis method was to learn. Since there was no tutorial or formal training with the analysis method in the study, there may have been little perceived learning regarding the method from the perspective of the participant. The lack of learning, however, need not have affected the ease of use associated with a method. After considering the need to include the ease of learning, the third "learning" item was dropped from the scale.

This left a two item scale including statements 1 and 2 in the list directly above. Table 10 below shows the alpha associated with all of the scales considered. The final "ease of use" scale used in the study is the two item scale. This scale used the items 1 and 2 described in the numbered list above.

Table 10: Reliability Analysis of the Scale Variable Ease of Use

Scale Description	Items used in Scale	Alpha Intergrammar	Alpha Intragrammar
4 –item Ease of Use	1, 2, 3, 4	0.605	0.554
3 –item Ease of Use	1, 2, 3	0.713	0.682
2 –item Ease of Use	1, 2	0.901	0.863

The loss of two of the items suggests that more attention needs to be given when adapting a previously successful scale to a new context for which the scale has not been designed.

Inter Rater Reliability: Problem Solving Questions

The problem solving tasks were described earlier in Chapter 4. Two individuals rated scores for the problem solving questions: the researcher and an independent rater[12]. The coding of the problem solving answers was "blind" for both individuals. This "blind" procedure was possible because the participant's packages were labeled with number that had no reference to the treatment group that the participant was from. The order of the packages was also randomized before coding so that the three treatment groups in each study were distributed evenly throughout the coding process. This effort was necessary in order to reduce the subjectivity of the problem solving ratings. Each rater was given a list of the acceptable answers to the five problem solving questions for each case. The problem solving questions and the list of acceptable answers to the problem solving questions are found in Appendix D.

When two individuals rate the same question that has an element of subjectivity, there are bound to be differences in the scores associated with some participants. To check the level of subjectivity the two ratings are normally tested through the use of a measure of correlation. Cohen's Kappa is used when the ratings are symmetric (i.e. the ratings have the same number of categories). The ratings for the problem solving questions were interval (at a minimum) and not symmetric. Cohen's Kappa, therefore, could not be used in this case. Instead, the interval nature of the data suggested the use of Pearson's correlation test statistic. Pearson's correlation is a number between zero and one that is a measure of the linear relationship between two variables. The closer the test statistic is to one, the stronger the linear association between the two variables. The two variables in this case are the two ratings provided by the independent raters. The resulting Pearson's correlation statistics along with the related levels of significance are shown below in Table 11.

Table 11: Inter Rater Reliability of Problem Solving Task

Case	Variables to Relate	Test Statistic	Value	Significance. (two tailed)
IFIP	IFPROBTOA IFPROBTOB	Pearson's ρ	0.904	0.000
Event	EVPROBTOA EVPROBTOB	Pearson's ρ	0.890	0.000
Voyager	VYPROBTOA VYPROBTOB	Pearson's ρ	0.919	0.000
Far East	FEPROBTOA FEPROBTOB	Pearson's ρ	0.898	0.000

The results in Table 11 provide a strong indication of the correlation between the problem solving scores for both raters. The high correlation suggests that the independent raters were largely in agreement in the assessment of the problem solving ability. In an effort to remain conservative, our estimates for problem solving score will make use of the rating provided by the external rater (**PROBTOB) rather than the rating provided by the researcher. The high correlation suggests that the external rating should serve as a useful proxy for the problem solving score.

The Relationship between Process and Effort Measures

While it is easy to see how the score on comprehension, problem solving, and Cloze tests are related to a construct of "understanding", it is more difficult to see how the times taken to complete these tests and the perceived "ease of use" are related to understanding. Elapsed times and ease of use are more closely related to a construct describing the "effort" with which understanding may be developed. The comprehension, problem solving and Cloze tests measure the **product** of understanding as concretized in the test scores. Elapsed times and ease–of-use are measures of the **effort** required in developing the understanding.

[12] I would like to thank Nanda Siva Kumar, PhD Student at the University of British Columbia for his hard work and efforts in coding the problem solving questions.

6.2 Preliminary Statistical Procedures for Intergrammar Study

The results of the main statistical procedures developed later in Chapter 7 are dependent on a number of preliminary statistical procedures. Preliminary statistical procedures for the intergrammar are discussed in this section. The preliminary statistics include: listing of descriptive statistics for demographic variables; analysis of demographics across treatment groups; analysis of dependent variables across treatment groups; analysis of case order on dependent variables; correlation between dependent and covariates; and tests for normality of dependent variables.

Demographics

The preliminary statistics begin with a listing of the descriptive statistics for all demographic variables collected in the analysis. These statistics provide some perspective on the study participants including their education, sex, and measures of their knowledge of the domain and analysis methods. An important preliminary analysis is to determine if there are any significant differences in the makeup of the participants across the three treatment groups (Text, DFD & ER, OOA). Table 12 below provides descriptive statistics for each of the dependent variables including the mean, minimum, maximum, and standard deviation across all 106 participants. Table 13 breaks the participants into the three treatment groups and provides means for variable across all three groups. The definitions for the variables included in the table are provided earlier in Table 7.

A one-way analysis of variance (ANOVA) procedure for each variable was used to test the null hypothesis that the means across the three treatment groups were equal. The resulting significance levels from these tests are provided in Column 4 of Table 13 below. The lower the significance level, the more likely that there is a difference between the treatment groups.

Table 12: Descriptive Statistics for Demographic Variables (106 participants)

Variable	Mean	Std Dev	Minimum	Maximum
EDU	.10	.31	.00	1.00
SEX	.42	.50	.00	1.00
PREEVENT	2.70	1.94	.00	7.00
PREIFIP	1.83	1.56	.00	7.00
PCHKEVEN	1.66	1.41	.00	5.00
PCHKIFIP	1.35	1.40	.00	5.00
PUSEDFD	.64	.48	.00	1.00
PUSEERD	.70	.46	.00	1.00
PUSEOOA	.53	.50	.00	1.00
PCMPDFD	3.00	2.15	.00	7.00
PCMPERD	3.53	2.05	.00	7.00
PCMPOOA	2.44	1.90	.00	6.00
PCNFDFD	2.80	2.05	.00	7.00
PCNFERD	3.34	2.02	.00	7.00
PCNFOOA	2.31	1.86	.00	7.00
PFAMDFD	2.94	4.63	.00	24.00
PFAMERD	3.66	3.87	.00	24.00
PFAMOOA	1.59	2.71	.00	24.00

Table 13: Comparison of Demographic Means Across Treatment Groups

Variable	Column 1 Text	Column 2 DFD/ERD	Column 3 OOA	Column 4 Significance (from F test in ANOVA)
# participants	**34**	**37**	**35**	
EDU	0.15	0.08	0.07	0.24
SEX	0.32	0.51	0.4	0.27
PREEVENT	2.38	2.78	2.91	0.49
PREIFIP	1.65	1.92	1.91	0.71
PCHKEVEN	1.50	1.57	1.91	0.42
PCHKIFIP	1.12	1.49	1.43	0.50
PUSEDFD	0.50	0.68	0.74	0.09
PUSEERD	0.56	0.73	0.80	0.08
PUSEOOA	0.47	0.54	0.57	0.69
PCMPDFD	2.68	2.97	3.34	0.43
PCMPERD	3.17	3.57	3.83	0.16
PCMPOOA	2.09	2.35	2.89	0.18
PCNFDFD	2.19	2.95	3.24	0.11
PCNFERD	2.61	3.59	3.77	0.07
PCNFOOA	1.96	2.38	2.57	0.13
PFAMDFD	2.76	2.51	3.53	0.19
PFAMERD	2.72	3.27	3.67	0.05
PFAMOOA	1.33	1.5	1.94	0.21

We can conclude, based on the results presented in Table 13 above, that there are no significant differences between the means of the three treatment groups for the demographic variables collected. Post hoc Scheffe tests were used to test for significant differences between two groups. These tests indicated that, taken as a single group, the "DFD/ERD and OOA" group was significantly different (at the (0.05 level) from the "Text" group on the variables PFAMERD and PCNFERD, and marginally significant on PCNFDFD. Upon further investigation, it was discovered that these differences could be attributed to a slightly higher proportion of participants who were quite knowledgeable about the modeling techniques in the DFD/ERD and OOA group. As a result, these individuals rated their confidence and experience with ERD and DFD much higher than the average participant.

While these differences are recognized, they do not suggest any systematic bias introduced during the random selection of participants into groups. Since the differences only occurred on some variables, I am reluctant to reduce sample size to eliminate these effects. Further, the 19 independent tests at the $\alpha = 0.05$ level suggest that level of accepted type I error for the entire test is equal to $(1- (0.95)^{19}) = 0.63$ which is high and suggests a high probability of some errant results. When a lower value of $\alpha = 0.01$ was used, no significant differences were observed and the cumulative type I error accepted is 18%. No adjustments were made, therefore, to the composition of the groups.

Testing for an Ordering Effect

The next step in the preliminary analysis is to test for an "ordering" effect across cases. The ordering effect refers to any learning effect that may have occurred due to the order in which the cases were presented. Participants were randomly distributed into six different versions of the test. Three of the versions started with the "Event" case and three of the version started with the "IFIP" case. We are interested in determining if there was any systematic difference when a case was used first as opposed to being used as the second case in the study.

To look for the ordering effect we will concentrate on the values of the dependent variables. If the ordering of the cases was important, the case order should affect the values of the dependent variables that were measured. To test for ordering effects, first we must split the data into the three groups that define the three presentation methods (Text, DFD & ER, OOA)). For each of these groups, we then identify two groups that identify the order that the cases were presented. We want to compare the mean scores in the first case ordering with the mean scores of the second case ordering To test for the ordering effect we use a one–way ANOVA for each dependent variable. Table 14 below shows the results of this analysis. Highlighted areas indicate significant differences.

Table 14: Test for Case Order Effects on Means of Dependent Variables (ANOVA)

	Column 1 ANOVA results Text Method			Column 2 ANOVA Results DFD/ERD Method			Column 3 ANOVA Results OAA Method		
Case Order	**IFIP then Event**	**Event then IFIP**	**Sig F test**	**IFIP then Event**	**Event then IFIP**	**Sig F test**	**IFIP then Event**	**Event then IFIP**	**Sig F test**
sample size	17	17	**34**	20	17	**37**	17	18	**35**
Variable									
IFCOMPTOT	8.52	8.52	.99	8.10	7.70	.41	7.41	7.22	.69
IFPRBTOA	5.88	6.4	.68	8.90	8.00	.56	9.76	11.5	.32
IFCLZTOT	18.3	25.7	.08	27.4	25.4	.41	25.29	25.16	.94
IFCMPTIM	9.28	9.79	.86	7.64	5.51	.40	5.82	6.95	.53
IFPRBTIM	13.70	15.56	.54	12.30	16.55	.11	16.23	16.55	.89
IFCLZTIM	10.05	9.32	.71	8.96	12.24	.04	10.92	12.00	.67
EVCOMPTOT	7.88	7.58	.62	9.70	9.35	.48	8.29	9.00	.21
EVPRBTOA	8.52	10.35	.35	6.95	5.70	.32	7.58	9.11	.32
EVCLZTOT	19.4	26.8	.06	22.15	21.82	.84	23.58	21.83	.35
EVCMPTIM	10.27	16.66	.02	12.22	13.57	.71	8.35	10.24	.48
EVPRBTIM	24.48	28.80	.14	20.24	20.27	.98	21.90	21.29	.78
EVCLZTIM	10.48	10.05	.78	12.27	12.57	.87	11.75	12.02	.88
EZTOUSE	6.42	6.64	.17	5.50	4.84	.10	5.05	4.16	.12
EZTOLERN	6.47	6.29	.41	5.30	4.72	.11	4.88	6.05	.48
EZFRUSRE	5.82	5.88	.90	5.20	5.05	.67	5.17	5.05	.64
EZUNDER	6.47	6.58	.55	5.5	4.7	.08	4.94	4.44	.34

From the results provided in Table 14, we can conclude that there is no systematic or significant difference in the dependent scores when the order of the cases were reversed. These results provide a justification for merging the data from both case orders into a single set of data, so that we can compare the Text, DFD & ER, and OOA methods without considering the effect of case order on the results.

Correlations Between Dependent Variables

One of the reasons for using MANCOVA is the assumption that the set of dependent variables are correlated and that separate ANCOVA analyses would not take this correlation into account. A correlation matrix between dependent variables can be used as a preliminary test of the assumption of dependence. A further test of dependent correlation is Bartlett's test of Sphericity. The result of Bartlett's test will be reported later in the analysis. Table 15 below shows the result of the correlation analysis on dependent variables for product. Results show Pearson correlation coefficient, sample size, and two- tailed significance level.

Table 15: Correlation Coefficients for Dependent Variables for Product

	Case: Event			**Case: IFIP**		
Variable	EVCMPTOT	EVPRBTOT	EVCLZTOT	IFCMPTOT	IFPRBTOB	IFCLZTOT
EVCMPTOT	1.00 p=.000	X	X	X	X	X
EVPRBTOT	-.099 p=.311	1.00 p=.000	X	X	X	X
EVCLZTOT	.084 p=.392	.2170 p=.025	1.00 p=.000	X	X	X
IFCMPTOT	**.268** **p=.006**	.001 p=.989	.191 p=.049	1.00 p=.000	X	X
IFPRBTOB	.136 p=.171	**.6264** **p=.000**	-.022 p=..817	.095 p=.329	1.00 p=.000	X
IFCLZTOT	.047 p=.628	.147 p=.132	**.649** **p=.000**	.129 p=.188	-.056 p=.565	1.00 p=.000

The results in Table 15 suggest a small degree of correlation between the dependent measures for product within each case. For example, there is a marginally significant correlation of $\rho = 0.21$ between EVPRBTOB and EVCLZTOT. What is more surprising are the correlations between cases. Note the extremely significant correlation between corresponding measures in the two cases (EVPRBTOB to IFPROBTOB = 0.62, and EVCLZTOT to IFCLOZTOT = 0.64, EVCMPTOT to IFCMPTOT = 0.27). These significant correlations provide some preliminary evidence that participants' scores are less case specific than previously thought. That is, participants who score above the mean on a dependent measure in one case are more likely to score above the mean on the same dependent measure used in another case.

The pattern of correlation also bodes well for the discussion of construct validity. The arrangement of correlations in Table 15 makes possible a technical evaluation of construct validity using the multi-trait, multi-method described in Straub (1989). This exercise will be left to Chapter 8 where validity is discussed.

Table 16 below repeats the correlation matrix this time for the variables relate to the effort required to interpret from the presentation method.

Table 16: Correlation Coefficients for Effort Variables

	Case: Event			**Case: IFIP**			**Ease**
Variable	EVCMPTIM	EVPRBTIM	EVZTIM	IFCMPTOT	IFPRBTOB	IFCLZTIM	EASE
EVCMPTIM	1.00 p=.000	X	X	X	X	X	X
EVPRBTIM	.157 p=.109	1.00 p=.000	X	X	X	X	X
EVCLZTIM	.137 p=.160	**.249 p=.010**	1.00 p=.000	X	X	X	X
IFCMPTIM	.127 p=.196	.109 p=.264	.064 p=.516	1.00 p=.000	X	X	X
IFPRBTIM	-.077 p=.430	**.298 p=.002**	.339 p=.000	-.019 p=.844	1.00 p=.000	X	X
IFCLZTIM	-.034 p=.728	.136 p=.164	**.519 p=.000**	-.055 p=..572	.156 p=.110	1.00 p=.000	X
EASE	**.272 p=.005**	**.198 p=.042**	-.106 p=..280	.042 p=.666	-.033 p=..735	-.114 p=.243	1.00 p=.000

(Pearson's Coefficient / p = two-tailed significance)

A similar pattern of correlation to those observed in Table 15 is observed among the dependent measures of process in Table 16. The correlations within variables in the same case indicate in general, there is little linear relationship between dependent variables.

Another interesting point is that the correlation between case variables, as observed in Table 15 above, is seen again in Table 16. These results are less surprising than those observed for product measures. High correlations between process variables in the same task, indicate a person that completed the test with below average time in the first case, will likely also be faster on the same task in a different case. Correlations between measures in a different task indicate that when a participant is quick at one of the tasks, they tend to be quick in other parts of the tests as well. For example the correlation between IFPRBTIM and EVPRBTIM of $\rho = 0.29$ indicates participants who are faster at the problem solving questions in the IFIP case are also faster in the problem solving questions in the Event case. As noted, more will be made of this late in Chapter 7.

Correlations between Dependent and Covariates

The final correlations we will look at in the preliminary analysis are the correlations between the dependent variables and the covariates we have developed for the study. These correlations are important as they provide initial evidence for either including or excluding the covariates in the MANCOVA analysis. Table 17 below presents the correlation matrix between all dependent measures and the covariates.

The results in Table 17 below indicate that the scale variables do not show consistent or significant linear relationship with the dependent variables used in the study. This provides us with an early indication that the covariates will have little effect in the MANCOVA analysis. More will be made of this in Chapter 7 when the MANCOVA analysis is discussed.

Table 17: Correlations between Dependent and Covariates

Dependent Variable	KNOWIFIP		KNOWEVEN		EASE	
	Corr.	Sig.	Corr.	Sig.	Corr.	Sig.
IFCOMPTOT	.004	.988	-.005	.957	-.084	.393
IFPRBTOB	.191	.050	-.013	.894	.079	.419
IFCLZTOT	-.190	.051	.041	.673	.011	.914
IFCMPTIM	-.093	.345	.197	.043	.070	.477
IFPRBTIM	-.080	.416	.197	.048	-.020	.840
IFCLZTIM	.102	.300	-.041	.675	.117	.230
EVCOMPTOT	.040	.684	-.070	.477	.081	.409
EVPRBTOB	.070	.476	-.080	.541	.046	.636
EVCLZTOT	-.240	.013	.047	.636	.043	.662
EVCMPTIM	-.050	.611	.013	.895	-.094	.336
EVPRBTIM	-.299	.002	.023	.814	-.057	.564
EVCLZTIM	-.023	.814	.130	.183	.141	.149
EZTOUSE	-.096	.329	-.064	.515	.079	.419

Normality of Dependent Variable Distributions

The final preliminary results reported in this section will be focused on the normality of the distributions of the dependent measures. The first assumption when using ANOVA and MANOVA procedures is that the dependent variable is distributed normally. As has been noted previously, ANOVA and large sample size MANOVA procedures are robust in relation to violations of the normality of the dependent variables. The test for homogenous covariances matrices (Box's M) is sensitive to departures from normality so the test for normality is important. Three tests will be employed to test the normality of the dependent variables: Wilks-Shapiro, skewness, and kurtosis. These tests are described in Appendix J. The results of these tests are shown below. Statistics that indicate a significant deviation from normal are highlighted in gray.

Table 18: Normality of Dependent Variables

Dependent Variable	TEXT GROUP W-S	Sig.	Skew.	Kurt	DFD & ER GROUP W-S	Sig.	Skew.	Kurt	OOA GROUP W-S	Sig.	Skew.	Kurt
IFCOMPTOT	.915	.016	.502	.164	.917	.013	.704	.295	.897	.010	1.08	2.75
IFPRBTOB	.977	.719	.310	-.15	.951	.184	.469	.195	.948	.166	.222	.948
IFCLZTOT	.909	.010	1.04	.791	.870	.010	1.54	2.93	.943	.097	.518	-.78
IFCMPTIM	.746	.010	2.06	4.14	.498	.010	4.56	9.85	.640	.010	2.82	8.38
IFPRBTIM	.933	.049	.488	.308	.835	.010	2.01	5.62	.953	.251	.735	.352
IFCLZTIM	.663	.010	2.30	4.70	.948	.141	.734	.268	.705	.010	3.03	11.7
EVCOMPTOT	.919	.020	.592	.275	.921	.017	.675	-.21	.920	.020	.185	.774
EVPRBTOB	.954	.262	.206	-.81	.947	.117	.236	.596	.910	.010	.628	.616
EVCLZTOT	.931	.045	.910	1.30	.878	.010	1.64	4.21	.974	.625	.280	-.21
EVCMPTIM	.759	.010	1.59	1.43	.741	.010	2.10	4.39	.538	.010	3.77	7.00
EVPRBTIM	.973	.609	.064	.631	.941	.076	.310	.959	.970	.524	.153	.852
EVCLZTIM	.934	.056	.403	-.78	.921	.017	.388	.759	.884	.010	1.21	1.00
EZTOUSE	.812	.010	.373	1.17	.936	.050	.962	2.49	.932	.046	.181	.326

Cutoff for skewness variable with $n = 30$ and $\alpha = 0.01$ is 1.114
Cutoff for Playkurtosis with $n = 30$ and $\alpha = 0.01$ is -1.21 and for Leptokurtosis is 2.21

The results in Table 18 above reveal that there are some dependent variables (particularly those associated with time to complete tasks) that are significantly different form normal. An important consideration, when examining Table 18, is how the distributions differ from normal. It should be noted that in our discussion in Appendix J, we noted that deviations from normality in regards to skewness are generally not considered serious if sample sizes are large. Deviations in kurtosis are serious when in the presence of platykurtosis (flattening of the distribution) but less serious (and more conservative) in the presence of leptokurtosis (stretching upwards of the distribution). A glance through Table 18 shows no evidence of significant platykurtosis. From this analysis we can conclude that the dependent variables for product are distributed normally, however, the dependent variables related to process show evidence of significant deviations from normal. Caution should be taken, therefore, in drawing references from the analysis of effort measures.

6.3 Preliminary Statistical Procedures for Intragrammar Study

The results of the statistical procedures related to the intragrammar study that are presented in Chapter 7 are dependent on a number of preliminary statistical procedures. These preliminary statistical procedures for the intergrammar study are discussed here. The preliminary statistics include: listing of descriptive statistics for demographic variables; analysis of demographics across treatment groups; analysis of dependent variables across treatment groups; analysis of case order on dependent variables; correlation analysis between dependent measures of product and process; correlation analysis between dependent and covariates, and tests for normality of dependent variable distributions. Later sections in Chapter 7 will reference these tests. A quick note for the reader is in order. The treatment described in this section for the intragrammar tests is very similar to the treatment provided to the intergrammar tests. If the reader has read section 6.3, then the reader has seen all of the preliminary techniques. The outcomes for both tests are summarized in Table 26 at the end of this chapter.

Demographics

The preliminary statistics for the intragrammar study begin with descriptive statistics for all demographic variables collected in the analysis. As noted earlier, these statistics provide some perspective on the study participants including their education, sex, and measures of their knowledge of the domain and analysis methods. An important preliminary analysis is to determine if there are any significant differences in the makeup of the participants across the three treatment groups (Text, DFD & ER, OOA). Table 19 below provides descriptive statistics for each of the dependent variables including the mean, minimum, maximum, and standard deviation across all 106 participants. Table 20 breaks the participants into the three treatment groups and provides means for variable across all three groups.

Table 19: Descriptive Statistics for Intragrammar Study (109 participants)

Variable	Minimum	Maximum	Mean	Std. Deviation
EDU	0.00	1.00	0.1357	0.501
SEX	0.00	1.00	0.4679	0.290
PREEVENT	0.00	6.00	1.789	1.427
PREIFIP	0.00	5.00	1.688	1.2451
PCHKEVEN	0.00	4.00	2.03	1.0620
PCHKIFIP	0.00	3.00	0.3486	0.6719
PUSEDFD	0.00	1.00	0.5138	0.5021
PUSEERD	0.00	1.00	0.6697	0.4725
PUSEOOA	0.00	1.00	0.3853	0.4889
PCMPDFD	0.00	24.00	2.862	4.296
PCMPERD	0.00	24.00	3.853	4.187
PCMPOOA	0.00	18.00	1.857	2.878
PCNFDFD	0.00	7.00	2.605	1.962
PCNFERD	0.00	7.00	3.827	1.994
PCNFOOA	0.00	7.00	2.330	1.986
PFAMDFD	0.00	7.00	2.651	1.926
PFAMERD	0.00	7.00	4.055	1.799
PFAMOOA	0.00	7.00	2.431	1.945

The statistics above provide an overview of the participants in the study, but does not provide an indication if participants in the three treatment groups were representative of randomized selection. A one-way analysis of variance (ANOVA) procedure for each variable was used to test the null hypothesis that the means for the demographic variables were equal across the three treatment groups. The resulting significance levels from these tests are provided in Column 4 of Table 20 below.

Table 20: Comparison of Demographic Means Across Treatment Groups

Variable	Column 1 Text	Column 2 ER Optional	Column 3 ER Mandatory	Column 4 Significance (from F test in ANOVA)
# participants	**32**	**38**	**39**	**109**
EDU	0.15	0.13	0.11	0.93
SEX	0.50	0.45	0.46	0.90
PROVOY	1.53	2.07	1.72	0.26
PROFAR	1.65	1.97	1.43	0.16
PCHKVOY	2.18	2.10	1.84	0.36
PCHKFAR	0.25	0.50	0.28	0.22
PUSEDFD	0.50	0.55	0.48	0.83
PUSEERD	0.59	0.73	0.67	0.46
PUSEOOA	0.37	0.47	0.31	0.33
PCMPDFD	2.34	2.97	2.46	0.35
PCMPERD	3.81	4.02	3.64	0.70
PCMPOOA	2.37	2.63	2.00	0.37
PCNFDFD	2.53	3.00	2.41	0.37
PCNFERD	4.09	4.34	3.74	0.34
PCNFOOA	2.40	2.97	1.92	0.06
PFAMDFD	2.56	2.98	2.98	0.89
PFAMERD	3.03	4.59	3.81	0.30
PFAMOOA	1.96	2.11	1.51	0.63

The results in Table 20 above indicate that there are no significant differences in demographic variables across the three treatment groups. Post hoc Scheffe tests were run to test the pairwise differences between the three treatment groups. None of the demographic variables showed significant results. These tests provide justification for the successful randomization of participants across the three groups.

Testing for an Ordering Effect

As in the previous intergrammar study, participants were randomly distributed into six different versions of the test. Three of the versions started with the "Voyager" case and three of the version started with the "Far East" case. This section addresses the "Case Order" question of whether there was a systematic difference when a case was used first as opposed to being used second in the study. To look for an ordering effect we split the data into the three groups that define the three presentation methods (Text, ER Optional, ER Mandatory). For each of these groups, we identify two orders in which the cases were presented. We compare the mean scores in the first case ordering with the mean scores of the second case using a one–way ANOVA for each dependent variable. Table 21 below shows the results of this analysis. Highlighted areas indicate significant differences at the α = 0.05 level. Note that the 39 independent tests below provide an overall $\alpha = (1-.95)^{39} = .87$ which is very high. A more acceptable overall alpha can be attained using α = 0.01 on each individual test. Under these conditions, none of the items are significant.

Table 21: Test for Case Order Effects on Means of Dependent Variables (ANOVA)

	Column 1 ANOVA results Text			**Column 2 ANOVA Results ER Optional**			**Column 3 ANOVA Results ER Mandatory**		
Case Order	**VOY then FAR**	**FAR then VOY**	**Sig F test**	**VOY then FAR**	**FAR then VOY**	**Sig F test**	**VOY then FAR**	**FAR then VOY**	**Sig F test**
Sample size	18	14	32	20	18	38	19	20	39
Variable									
VBCOMPTOT	6.55	7.57	.11	6.15	7.00	.06	6.94	6.85	.84
VBPRBTOB	13.88	11.71	.17	10.95	11.66	.60	13.47	12.95	.64
VBCLZTOT	31.39	32.28	.62	26.9	23.33	.08	31.42	26.71	.01
VBCMPTIM	10.31	9.95	.91	9.05	10.09	.56	12.05	9.12	.30
VBPRBTIM	19.46	19.10	.89	14.88	18.90	.09	18.08	14.85	.24
VBCLZTIM	9.01	8.71	.84	9.51	10.69	.32	9.85	8.96	.34
FECOMPTOT	8.27	7.92	.70	6.25	7.50	.06	7.36	7.00	.62
FEPRBTOB	13.39	12.5	.61	12.30	14.33	.05	12.78	14.25	.16
FECLZTOT	32.89	33.50	.77	27.20	24.66	.18	30.26	28.55	.29
FECMPTIM	15.34	9.37	.20	11.92	14.57	.23	11.91	13.47	.63
FEPRBTIM	21.65	28.74	.02	19.82	21.87	.08	18.37	22.04	.16
FECLZTIM	7.92	10.32	.12	9.25	12.57	.03	10.20	10.15	.97
EASE	1.76	2.40	.49	-0.76	-.02	.28	-1.19	-1.34	.81

We can conclude from the results provided in Table 21 that there is no systematic or significant difference in the dependent scores when the order of the cases were reversed. These results justify the merging of both case orders into a single set of data. We can now justify the comparison of the Text, ER Optional, and ER Mandatory methods without considering the effect of case order on the results.

Correlations Between Dependent Variables

As described above a correlation matrix between dependent variables can be used as a preliminary test of the assumption of dependence between dependent variables. This dependence provides a justification for using a single multivariate technique rather than multiple univariate techniques. Table 22 below shows the result of the correlation analysis on dependent variables for product. This table is followed by Table 23, which shows the result of the correlation analysis on dependent variables for process. Results in these two tables show the Pearson correlation coefficient and two tailed significance level. Significant correlations (α = 0.01) are highlighted in bold.

Table 22: Correlation Coefficients for Dependent Variables for Product

	Case: Voyager			Case: Far East		
Variable	EVCMPTOT	EVCMPTOT	EVCLZTOT	IFCMPTOT	IFPRBTOB	IFCLZTOT
EVCMPTOT	1.00 p=.000	X	X	X	X	X
EVCMPTOT	.005 p=.955	1.00 p=.000	X	X	X	X
EVCLZTOT	.001 p=.995	..039 p=.686	1.00 p=.000	X	X	X
IFCMPTOT	**.391 p=.000**	.178 p=.064	.118 p=..221	1.00 p=.000	X	X
IFPRBTOB	-.069 p=.479	**.504 p=.000**	-.099 p=.308	.020 p=.836	1.00 p=.000	X
IFCLZTOT	.117 p=.226	.16 p=.083	**.708 p=.000**	.188 p=.050	-.067 p=.491	1.00 p=.000

The results in Table 22 suggest very little correlation between the dependent measures for product within each case. This suggests that the dependent measures could be treated in separate ANOVA analysis. A more powerful procedure Bartlett's sphericity will be described in the MANCOVA procedure in Chapter 7. Our analysis will continue to use MANOVA as we can handle all significance tests in a single procedure. This will reduce the level of alpha for the overall test.

Again, the high correlation between the method of test (for example comprehension scores in Voyager case with comprehension scores in Far East Case) is remarkable. Similar correlation results were also observed in the other study. The consistent and surprising strength of the correlations between dependent measures across cases requires more detailed analysis that will follow in Chapter 7.Continuing with the correlation between dependent measures, Table 23 below repeats the correlation matrix shown in Table 26 for the dependent variables related to process.

Table 23: Correlation Coefficients for Effort

	Case: Voyager			Case: Far East			Ease
Variable	**EVCMPTOT**	**EVPRBTOT**	**EVCLZTOT**	**IFCMPTOT**	**IFPRBTOB**	**IFCLZTIM**	**EASE**
FECMPTIM	1.00 p=.000	X	X	X	X	X	X
FEPRBTIM	.033 p=.734	1.00 p=.000	X	X	X	X	X
FECLZTIM	**.256 p=.007**	**.379 p=.000**	1.00 p=.000	X	X	X	X
VBCMPTIM	.027 p=.782	-.018 p=.853	-.045 p=.635	1.00 p=.000	X	X	X
VBPRBTIM	.160 p=.096	**.305 p=.001**	.197 p=.040	.066 p=.495	1.00 p=.000	X	X
VBCLZTIM	.149 p=.122	.205 p=.033	**.602 p=.000**	-.061 p=.529	.242 p=.011	1.00 p=.000	X
EASE	.065 p=.499	.057 p=.533	-.206 p=.031	.163 p=.091	.141 p=.143	-.153 p=.113	1.00 p=.000

(Pearson's Coefficient / p = two-tailed significance)

The results in Table 23 suggest that there is stronger evidence of correlation between dependent measures (within each case). This evidence can be used to suggest that a multivariate technique should be used to analyze the process measures. Another interesting point is that the correlation between case variables, as observed in Table 15, 16, and 22 above, are again seen again in Table 23. As noted earlier, these results are less surprising than those in Table 22. Significant correlations between process variables in the same task (for example comprehension), indicate a person that completed the test with below average time in the first case, will likely also be faster on the same task in a different case. For example, the correlation between VBPRBTIM and FEPRBTIM of ρ = 0.305 indicates participants who are faster at the problem solving questions in the "Voyager" case are also faster in the problem solving questions in the "Far East" case. Correlations between measures in a different task indicate that when a participant is quick at one of the tasks associated with a case (for example the Cloze test), they tend to be quick in other parts of the tests as well (such as comprehension or problem solving). As noted, more will be made of these correlations in Chapter 7.

Correlations between Dependent and Covariates

In this section we focus on the correlations between the dependent variables and the covariates for the intragrammar study. These correlations are important as they provide initial evidence for either including or excluding the covariates in the MANCOVA analysis. Table 24 below presents the correlation matrix between all dependent measures and the covariates.

Table 24: Correlations between Dependent and Covariates

Dependent Variable	KNOWVOY Corr.	KNOWVOY Sig.	KNOWFAR Corr.	KNOWFAR Sig.	KNOWMETH Corr.	KNOWMETH Sig.
VBOMPTOT	-.127	.188	.024	.802	.042	.664
VBRBTOB	.026	.729	-.049	.616	-.022	.624
VBLZTOT	-.073	.448	-.008	.933	-.065	.501
VBCMPTIM	.062	.521	.024	.808	-.063	.517
VBPRBTIM	.191	.046	.046	.636	-.181	.060
VBCLZTIM	.228	.017	.074	.447	-.053	.584
FECOMPTOT	-.003	.997	.052	.595	,003	.971
FEPRBTOB	.052	.595	.032	.738	.082	.396
FECLZTOT	-.093	.338	.019	.841	.002	.987
FECMPTIM	-.128	.183	.060	.539	.114	.239
FEPRBTIM	.231	.016	.199	.038	-.157	.102
FECLZTIM	.152	.558	-.090	.385	-.018	.850
EZTOUSE	-.057	.558	-.090	.353	.082	.397

The results in Table 24 above indicate that the scale variables show no consistent and significant linear relationship with the dependent variables used in the study. This provides us with an early indication that the covariates will have little effect in the MANCOVA analysis. More will be made of this in Chapter 7 when the MANCOVA analysis is discussed.

Normality of Dependent Variable Distributions

This section reports on the normality of the dependent variable distributions. These tests are necessary to establish normality for test of homogenous covariance matrices (Box's M). Three tests are used to test the normality of the dependent variables: Wilks-Shapiro, skewness, and kurtosis. These tests have been described in Appendix J. The results of these tests are shown below in Table 25. Significant deviations from the normal distribution are highlighted in gray.

Table 25: Normality of Dependent Variables

Dependent Variable	TEXT GROUP W-S	Sig.	Skew.	Kurt	ER MANDATORY W-S	Sig.	Skew.	Kurt	ER OPTIONAL W-S	Sig.	Skew.	Kurt
VBCOMPTOT	.948	.201	-.66	.831	.922	.017	-.21	-.86	.878	.010	-.95	1.42
VBPRBTOB	.931	.055	.653	-.29	.976	.866	.172	-.46	.963	.361	.494	.291
VBCLZTOT	.946	.169	-.06	-.66	.982	.829	-.18	-.16	.946	.087	-.80	1.22
VBCMPTIM	.758	.010	2.10	5.09	.634	.010	2.45	5.05	.697	.010	2.43	6.24
VBPRBTIM	.909	.014	.943	1.06	.929	.029	.821	.092	.948	.108	.483	-.50
VBCLZTIM	.812	.010	2.00	5.12	.945	.089	.934	1.25	.947	.096	.432	-.19
FECOMPTOT	.865	.010	-1.2	1.30	.907	.010	-1.2	3.20	.951	.147	-.38	-.54
FEPRBTOB	.967	.483	-.13	-.64	.968	.447	-.24	.736	.958	.278	.287	-.23
FECLZTOT	.968	.501	-.02	0.88	.962	.374	-.49	0.17	.954	.204	-.69	.526
FECMPTIM	.676	.010	2.08	3.60	.708	.010	3.67	8.95	.741	.010	1.82	2.52
FEPRBTIM	.959	.364	.494	-.02	.880	.010	1.07	.423	.943	.072	.463	-.77
FECLZTIM	.830	.010	1.14	.121	.956	.248	.751	1.32	.849	.010	1.62	2.91
EZTOUSE	.924	.038	-.61	-.12	.978	.728	-.07	.691	.948	.101	.537	-.25

Cutoff for skewness variable with $n = 30$ and $\alpha = 0.01$ is 1.114
Cutoff for Playkurtosis with $n = 30$ and $\alpha = 0.01$ is -1.21 and for Leptokurtosis is 2.21

The results in Table 25 above reveal that there are some dependent variables (particularly those associated with time to complete tasks) that are significantly different from normal. When compared to Table 18 for the intergrammar study, there are less distributions that significantly differ from normal in the intragrammar study than in the intergrammar study.

An important consideration, when examining Table 25, is how the distributions differ from normal. We observe again that most deviations from normal are the result of leptokurtosis which is less harmful to the power of the test that platykurtosis. From this analysis we can conclude that the dependent variables for product are distributed normally, and that most of the dependent variables related to process (except for times to complete comprehension tasks) are not significantly and consistently different from normal. Caution should still be taken, however, when inferring from the analysis of process measures.

6.4 Summary

The intent in writing this chapter has been to reduce the volume of statistical tests surrounding the main hypotheses in this thesis. The reader may be somewhat disheartened, however, by the fact that the myriad of results presented in the chapter represent only the beginning of the analysis. In truth, the heart of the matter has been revealed in the pieces of evidence strewn throughout this chapter. A close look at Tables 14 and 21, for example, provides enticing initial evidence of difference brewing across treatment groups. One can also see themes running through each of the fours cases used in this thesis. While the researcher recognizes the daunting realization that the reader must endure yet another chapter of hypothesis tests, I can promise that a good foundation has been laid, and an interesting story will emerge.

The preliminary tests developed and reported in this chapter are elements in a thorough analysis of the data collected. It is hoped that this preliminary analysis has provided a solid foundation for the analysis that follows in Chapter 7. Out of respect to the reader, Table 26 is offered on the next page as a summary of all of the findings in this chapter. This table will be referenced in the following chapter.

Table 26: Summary of Chapter 6 Results

Item	Study: Intergrammar		Study: Intragrammar	
Scale Variable Reliability				
	Case	**Cronbach's α**	**Case**	**Cronbach's α**
Domain Knowledge	IFIP	0.737	Voyager	0.438
	Event	0.764	Far East	0.533
Knowledge of Analysis Methods	IFIP & Event	0.872	Voyager & Far East	0.861
Ease of Use	IFIP & Event	0.901	Voyager & Far East	0.861
Inter Rater Reliability				
	Case	**Pearson's ρ**	**Case**	**Pearson's ρ**
Inter-rater Reliability for Problem Solving Questions	IFIP	0.904	Voyager	0.919
	Event	0.890	Far East	0.898
Preliminary Questions				
	Case	**Answer**	**Case**	**Answer**
1. Is there evidence that the process and product measures are independent?	IFIP	Yes	Voyager	Yes
	Event	Yes	Far East	Yes
	Case	**Answer**	**Case**	**Answer**
2. Are there significant differences on pre-test variables between the treatment groups?	IFIP	No	Voyager	No
	Event	No	Far East	No
	Case	**Answer**	**Case**	**Answer**
3. Is there a significant ordering effect so that scores change if a case is presented second rather than first in the study?	IFIP	No	Voyager	No
	Event	No	Far East	No
	Case	**Answer**	**Case**	**Answer**
4. Is there evidence to support the idea that there is correlation between the dependent variables?	IFIP	Yes	Voyager	Yes
	Event	Yes	Far East	Yes
	Case	**Answer**	**Case**	**Answer**
5 Is there evidence that the covariates are correlated with dependent variables?	IFIP	No	Voyager	No
	Event	No	Far East	No
	Case	**Answer**	**Case**	**Answer**
6. Is there evidence to support the statement that the dependent variables are distributed normally?	IFIP	Somewhat	Voyager	Somewhat
	Event	Somewhat	Far East	Somewhat

Chapter 7: Results and Discussion

This chapter will report on the results of the intergrammar and intragrammar comparisons discussed earlier in chapters 4, 5 and 6. The discussion will be organized in relation to the four hypotheses developed in Chapter 4 and summarized in Table 3. Table 3 is recreated below to provide a quick overview of the finding in the two studies. While much will be said throughout this chapter regarding the statistical procedures and tests for significance, Table 27 is the heart of the matter. The predictions that are substantiated in this chapter are highlighted in gray. Those predictions that could not be substantiated remain in white. The remainder of the chapter discusses how the assertions made in Table 27 were arrived at.

Table 27: Summary of Research Findings

		Hypotheses					
	Type	**Measure**		**H1.1** p. 56	**H1.2** p. 60	**P2.1** p. 67	**P3.1** p.70
Predictions	Product	Comprehension		Graphic> text	OOA => DFD/ERD	Mand = Opt	HPKM* > LPKM**
		Problem Solving		Graphic > text	OOA > DFD/ERD	Mand > Opt	HPKM > LPKM
		Text Reconstruction (Cloze test)		Graphic < text	OOA > DFD/ERD	Mand > Opt	HPKM > LPKM
	Process	Time to Complete Task	Comp.	Diagram < text	OOA =< DFD/ERD	Mand =< Opt	HPKM < LPKM
			Prob.	Diagram < text	OOA =< DFD/ERD	Mand =< Opt	HPKM < LPKM
			Cloze	Diagram < text	OOA =< DFD/ERD	Mand =< Opt	HPKM < LPKM
		Perceived Ease of Use		Graph > text	OOA => DFD/ERD	Mand = Opt	HPKM > LPKM

<u>Abbreviations used in Table 27</u>

Graph	- graphical description	Text	- text description
OOA	- Object Oriented Analysis	DFD/ERD	– Data flow and ER diagram
Mand	- mandatory with sub-typing	Opt	- Optional attributes
*HPKM	– high previous knowledge of method		
*LPKM	– low previous knowledge of method		

In Table 27 above, we see that many predictions made in Chapter 4 are substantiated by the empirical tests, particularly those regarding the product measures. Other predictions, primarily related to the process measures, are not. The remainder of this chapter is concerned with the explanation of how these assertions were arrived at, and to begin to discuss why these results were observed. The discussion begins with the choice of the statistical procedure.

Why Use MANOVA?

If the dependent variables used in a MANOVA were independent of each other, one could run a series of univariate ANOVA procedures (one for each dependent variable) and come to the same conclusions. The use of MANOVA must be justified, therefore, before the procedure is used in this thesis. The MANOVA procedure was chosen for this study, over more traditional ANOVA techniques, for the following four reasons adapted from Stevens (1992).

1. The treatment variable, (presentation method), will affect the participant's understanding of the domain in a complex way. To understand the effect of the treatment variable we will have to measure more than one aspect of "understanding". This suggests that we collect and analyze ***multiple*** measures of a latent dependent variable that we label as "understanding". An ANOVA procedure that analyzes only a single dependent measure is not appropriate.

2. These multiple measures of the construct "understanding" share a common "meaning" and should not be treated separately. It is important to distinguish between the latent variable of "understanding" and the concrete realization of "understanding" represented by the three dependent measures (comprehension, problem solving, and understanding). MANOVA is a technique for addressing the analysis of latent variables using multiple measures.

3. The use of fragmented univariate ANOVA tests, instead of a single MANOVA procedure, can lead to inflated overall type I error (false rejection). For example, consider a construct measured with 4 dependent variables. If α=0.05 for each of the 4 univariate tests, the probability that we have at least one false rejection is approximately equal to $(1 - (.0.95)^4)$ or 18%, which is unacceptably high.

4. The treatment groups may not be significantly different on any of the dependent variables individually, but when considered jointly the dependent variables may be differentiated reliably across the treatment groups. Multivariate tests are more powerful than univariate techniques in these types of cases.

MANOVA is a statistical technique designed to compare groups of subjects across several dependent measures, where the dependent variables are correlated and share a common conceptual meaning. The numerous dependent measures collected in this study are assumed to be correlated and share the common conceptual. This fact, combined with the five reasons provided above, suggests that MANOVA is an appropriate choice for a statistical analysis technique in this study.

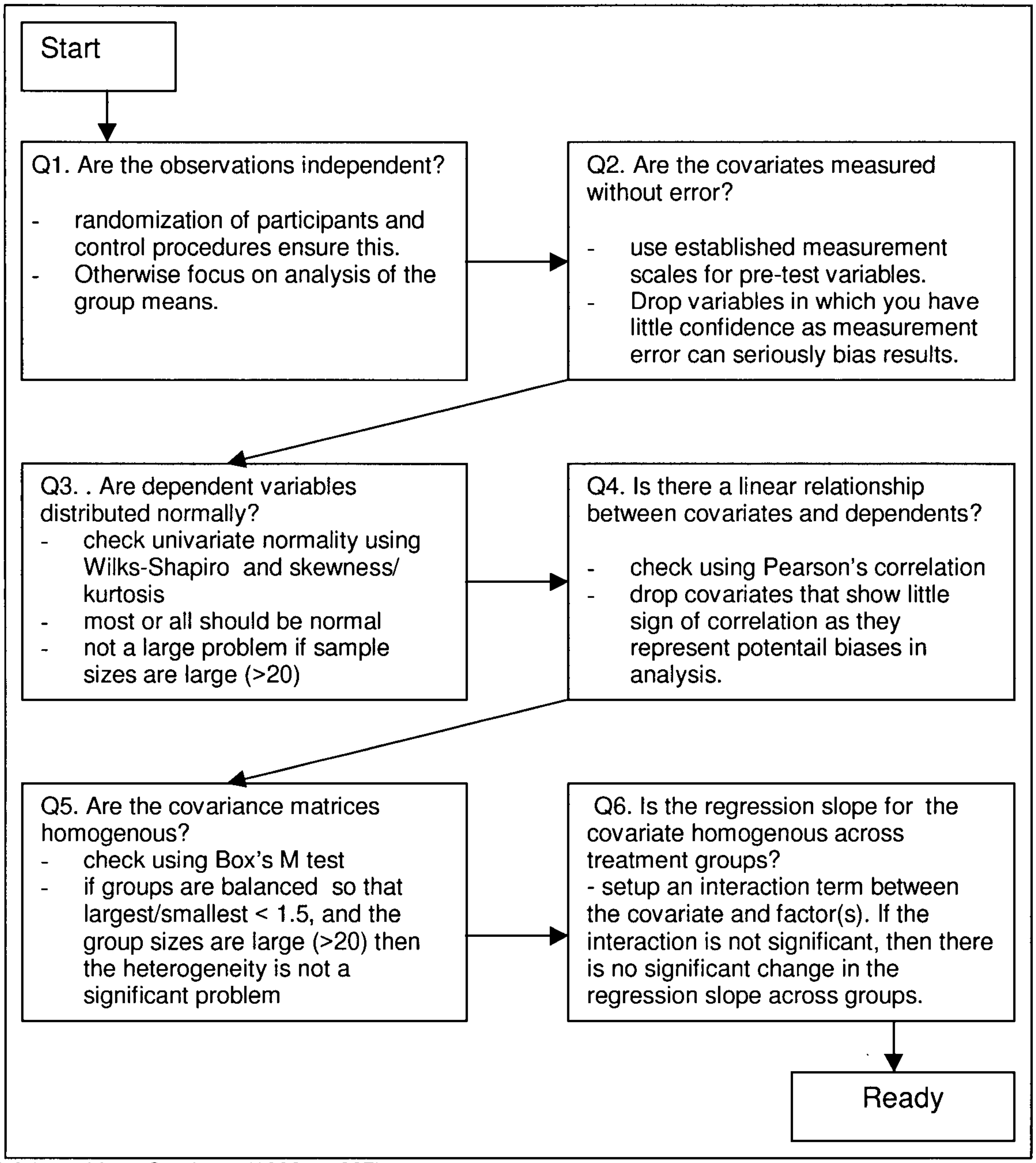
7.1 Procedure for Assessing Assumptions in MANCOVA

The assumptions that underlie the MANOVA and MANCOVA procedures are outlined in Appendix J. Before hypothesis tests can begin for any MANCOVA model, these assumptions must be verified. To increase the clarity and reduce the complexity associated with the assessment of the MANCOVA assumptions, Figure 10 on the next page provides a simple flow chart for the assessment process. This diagram builds on the flow chart provided by Stevens (1991, p. 267). For a more detailed description of the assumptions and the consequences of violating the assumptions see Appendix J and Stephens (1992) Chapter 6, 9.

The flow chart in Figure 10 provides us with a step by step procedure for assessing the assumptions before developing our analysis. Note that the first two steps in the flow chart do not require any statistical analysis. These steps are covered in the initial design of the experiment and the pretest measures used in the analysis. The experimental design used in this study enables us to states that observations are independent. We have also noted, in Section 6.1, that of the two covariates available in the study (knowledge of method and knowledge of domain), the scale variables for the knowledge of the domain (KNOWIFIP, KNOWEVEN, KNOWVOY, KNOWFAR) are relatively weak. It will be difficult to find reason to include these scale variables in the final analysis. That is, the measures of domain knowledge do not adequately capture the knowledge of the domain.

The next two steps in the assessment procedure, steps three and four, have been completed in the preliminary analysis. These steps address whether the dependent measures are distributed normally and whether there is a significant relationship between covariates and dependent variables. Our preliminary results suggest that in both studies, the dependent variable populations could be considered normally distributed, and that none of the covariates used in the study shared a consistent or significantly large correlation with any of the dependent variables. These finding should be treated carefully for two reasons. First, we do not have a test for multivariate normality. So this assumption cannot be addressed directly. Second, we must rely on univariate correlation measures (Pearson's coefficient) for estimates of a linear relationship between covariates and dependent variables. Despite these limitations, the steps are important factors for the final analysis and are included here.

Figure 10: Flow Chart of the Assessment Process *



* Adapted from Stephens (1992, p. 267)

The final two steps in the assessment process are tests that require the MANCOVA procedure to be run and are part of the output provided by the MANOVA procedure in SPSS. Note that the experimental design described in Chapter 5 contributed in reducing the impact of violations in the assumptions in utilizing a balanced design and a large (>30) group size.

Figure 10 above outlines a process that must be performed for each MANCOVA procedure. There are four tests that make up the assessment process. These are:

1. **Check of univariate normality test for dependent variables**
2. **Check correlation matrix of dependent variables and covariates**
3. **Check Box's M test for homogenous covariance matrices**
4. **Check factor *covariate interaction for significance.**

The first two items have already been completed for both studies, and items three and four are contained in MANCOVA output. This analysis of the assessment of the assumptions leads us next to the procedures used for analyzing MANCOVA results. This process is described below.

7.2 Procedure for Analyzing MANCOVA Results

In this section we focus on how the results provided in the MANCOVA will be analyzed. Our preliminary analysis has enabled us to state three things. First, the distributions of dependent measures do not systematically or consistently deviate from the normal distribution. This is true for all cases used in this study, and particularly true for the product measures. Second, correlations between dependent variables and covariates were largely insignificant across all cases. This provides preliminary indications that covariates will not be significant factors in the MANCOVA. Finally, two additional assumptions need to be addressed. These assumptions are the assumption of homogeneous dependent measure covariance matrices, and homogeneous regression slopes across treatment groups. Any analysis of the MANCOVA requires that we address these two additional assessments before making inferences.

Once the assumptions have been addressed the analysis of the results can begin. There are a number of results of interest in the study. Rather than enumerating all possible tests, I have decided to focus my attention on the tests provided in the list below. The list provides two types of information. First, it provides the list of tests used in the analysis of the MANCOVA results. Second, it provides the order in which these tests are performed. The order is important since, in some cases, the MANCOVA may have to be re-estimated as a result of one of the test. The list contains the following items (along with an explanation of each item):

1. **Bartlett's Test of Sphericity**. This statistic is used to test whether the dependent variables are correlated. The null hypothesis is that the dependent correlation matrix is equal to an identity matrix which suggests that diagonal element are one and off diagonal elements are zero. If this null hypothesis is not rejected, MANOVA may not be appropriate and separate ANOVA procedures are recommended.

2. **Box's M Test**. This statistic is used to test whether the covariance matrices are equal (homogeneous). The effects of heterogeneous covariances can be large but are

moderated by equal groups size and a large sample size. This test is sensitive to the underlying normality of the dependent variable distributions.

3. **Significant Within Cell Regression Effect**. This test checks for a significant linear relationship between the covariates and the dependent variable. If there is no significant effect, then the covariate is not appropriate and further analysis should be done with the effects of the covariate removed. This would require re-estimating the MANOVA without the covariate.

4. **Test for Multivariate Differences (Interaction Effects).** In cases where there is more than one treatment variable or where there is a treatment*covariate interaction term included in the model, the interaction effect should be tested before looking to the main effects. If the treatment*covariate effect is significant, the assumption of homogenous regression slopes is violated. Also, a significant interaction effect will increase the difficulty of interpreting the main effects, and the main effects may be misleading in the presence of a significant interaction effect.

5. **Test for Multivariate Differences (Main Effects)**. This is the main statistic developed in MANOVA. It indicates whether a significant difference exists across treatment groups for the set of dependent variables. Four measures are provided by SPSS (Pillai's trace, Wilk's lamda, Hotelling's trace, and Roy's largest root). Stephens (1992, p. 268) notes "*For variance differences likely to occur in practice... Pillai-Bartlett trace, Wilk's lamda, and Hotellings-Lawley trace are equally robust.*" The test for multivariate differences will make use of Wilk's lamda as described earlier in section 6.1.

6. **Tests for Univariate Differences**. When multivariate difference are found to be significant, the related univariate tests can be used to identify which dependent variables contribute to the overall differences. The univariate tests are equivalent to One-way ANOVA for each dependent variable. These post hoc tests are essential for determining the source of the differences between treatment groups.

The set of six tests described above, along with the assessment procedure described in the previous section, combine to make up the overall testing procedure for every MANCOVA analysis described in this chapter. It is hoped that outlining the procedures, as we have in these two sections, will help to improve the clarity of the analysis. With these descriptions out of the way, it is time to move on to the analysis of the four hypothesis summarized in Table 27 above.

7.3 Hypothesis H1.1: Text vs. Diagrams

We begin the analysis of the predictions made for Hypothesis H1.1 by restating the hypothesis:

> **H1.1: Diagrams will be better able to promote understanding in individuals interpreting a domain description than text .**

From this hypothesis we predicted that:

1. Comprehension and Problem Solving scores will be higher, on average, for participants with diagrams than text descriptions.
2. Cloze score will be higher, on average, for participants with text descriptions because these participants had been exposed to the text prior to the Cloze test whereas other participants had not been exposed.
3. Participants viewing diagrams will take less time on the comprehension and problem solving, and may take more time on the Cloze tasks, on average, than participants provided with text.
4. Participants will perceive diagrams to be easier to use than text descriptions to complete study tasks.

To test these predictions we create MANCOVA analyses using the data for the intergrammar study. The intergrammar study has three treatment groups: Text, DFD & ER, and OOA. The OOA and DFD & ER represent graphical diagrams. Two covariates - KNOWMETH and KNOW**** (where **** is either IFIP or EVEN depending on the case) - are initially used. The comprehension, problem solving, and Cloze score were dependent variables. The effort variables are also included in the analysis. These effort variables include the time taken to complete each of the tasks as well as the perceived "ease of use" scale variable. Two cases were used to further the external validity associated with the findings, so two separate MANCOVA analyses associated with hypothesis H1.1 were estimated. We begin the discussion of the results with a look at the means and standard deviations. This is shown below in Table 28.

In developing and analyzing comparisons between the participant groups, it is important to have a view of the where the comparisons start. Table 28 contains the means and standard deviations for each of the three groups considered in the study for all of the dependent measures in the study (including product and effort). The table makes no claims for significant differences and no statistical transformations are applied to the data. The numbers are simply averages and associated standard deviations. Readers are encouraged to look over this table, as these numbers form the basis of almost all of the claims made for the first two hypotheses.

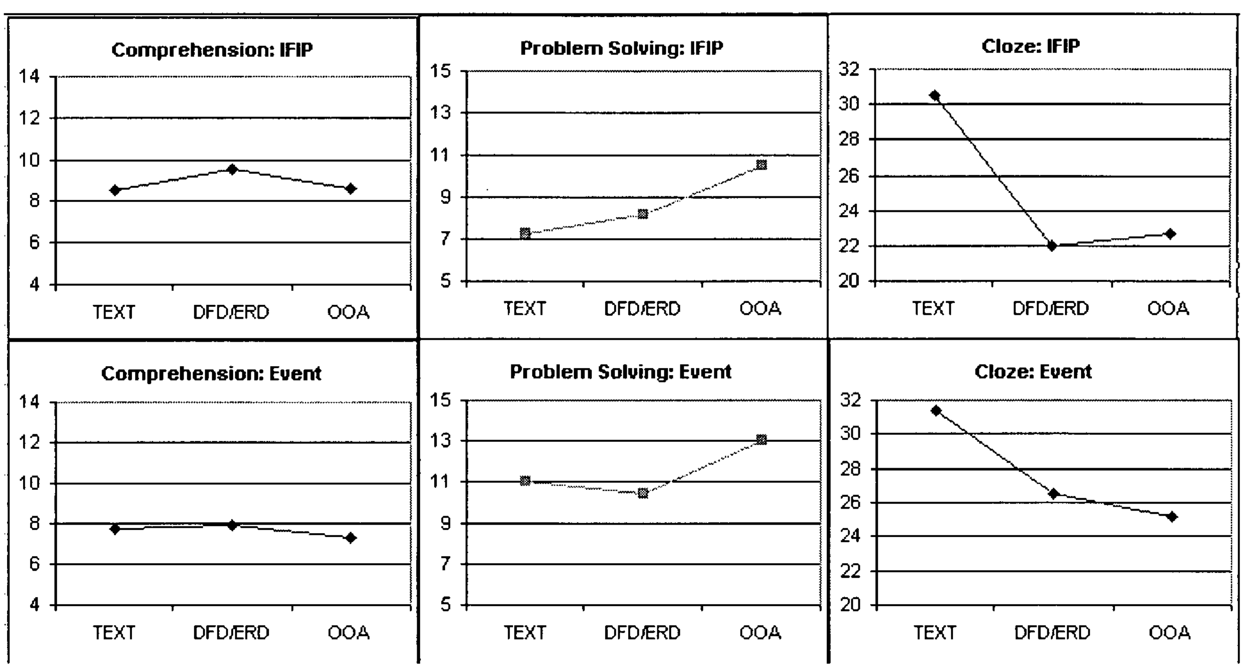
Table 28: Means and Std. Dev. Across Treatment Groups for Dependent Variables

Dependent Measure	Case: IFIP			Case: EVENT		
	TEXT	DFD/ERD	OOA	TEXT	DFD/ERD	OOA
COMP	8.52 (1.65)	9.54 (1.48)	8.65 (1.66)	7.73 (1.69)	7.91 (1,44)	7.31 (1.51)
PROB	7.26 (2.79)	8.18 (2.60)	10.54 (4.20)	11.02 (3.68)	10.40 (2.26)	13.05 (4.21)
CLOZ	30.47 (8.32)	22.00 (4.91)	22.68 (5.47)	31.41 (8.47)	26.48 (7.20)	25.22 (5.26)
COMP TIME	9.54 (8.70)	6.65 (7.48)	6.40 (5.17)	13.68 (10.45)	12.84 (10.73)	9.32 (10.06)
PROB TIME	14.62 (8.66)	14.25 (8.20)	16.38 (6.61)	23.63 (8.50)	20.25 (7.62)	21.59 (6.31)
CLOZ TIME	9.69 (5.70)	10.47 (4.77)	11.49 (7.35)	10.27 (4.08)	12.41 (5.68)	11.89 (5.07)
EASE	1.967 (.832)	-0.431 (1.71)	-1.44 (1.11)	1.85 (.812)	-.427 (1.58)	-1.35 (1.31)

* Numbers in brackets are standard deviations

Having looked at the means, we can quickly compare them using graphs. To save space, only the means for the comprehension, problem solving, and Cloze scores for the three treatment groups across two cases are provided below in Figure 11.

<u>Figure 11: Means for Comprehension and Problem Solving: IFIP and Event Case</u>



The graphs show some evidence of differences. To test for significant difference we will begin with the multivariate approach and then move to the univariate analyses. The results from the two MANCOVA analyses are provided below in Table 29.

Table 29: MANCOVA for Hypothesis H1.1

Item	Case: IFIP		Case: Event	
1. Bartlett's Sphericity				
	Chi	**Sig.**	**Chi**	**Sig.**
Are dependent Var.correlated?	26.88	.012	29.75	.005
2. Box's M				
	M	**Sig.**	**M**	**Sig.**
Homogenous covar. matrices?	93.27	.009	122.5	.000
3. Within Cell Regression Effect				
Covariates significant?	λ	**Sig.**	λ	**Sig.**
KnowMeth	.007	.872	.035	.317
KnowIFIP	.003	.992	N/A	N/A
KnowEVEN	N/A	N/A	.019	.596
4. Factor*covariate Interaction Effect				
	λ	**Sig.**	λ	**Sig.**
Homogenous regress. effects?	N/A	N/A	N/A	N/A
5. Main Effect				
	λ	**Sig.**	λ	**Sig.**
Significant effects between methods?	.676	.000	.716	.000

Table 29 above provides us with a large amount of information. First, the Bartlett tests of sphericity indicate that the MANOVA procedure is appropriate as the dependent variables in all cases show a significant amount of multivariate correlation. Second, the Box's M test indicates that the assumption of homogeneous covariance matrices is violated in all of the procedures. As discussed in Appendix J, the large sample size and relatively equal group sizes in this study reduce the impact of the violation of the assumption. The effect of the violation is to decrease the power of the test. Type 1error rates are affected only slightly except in extreme cases, and there is nothing to indicate that the conditions in this case are extreme.

The covariates in both MANCOVA procedures were not significant. This finding was to be expected given the very low correlations observed between dependent measures and the covariates. Since the covariates were not significant, the main effects could be estimated without including the effect of the covariates. The covariates were therefore dropped from all four of the models. Dropping the covariates eliminated the need for the factor*covariate interaction effect, which was not reported.

The main effects show strong significance in both the MANCOVA procedures. Wilks λ(s=2, m=0, n=48) = .676 at p <.001 for the IFP case and λ (s=2, m=0, n=48) = 0.716 at p < .001 for the event case. These findings suggest that there are significant differences among the treatment groups in both cases. While these finding are encouraging, it is necessary to perform a univariate post hoc analysis on the dependent variables to see which variables contributed to the observed differences.

The univariate tests are one-way ANOVA procedures that produce an F statistic. The univariate tests provide a method for addressing the predictions made in Table 27 as a result of our Hypothesis. These univariate tests are, therefore, an important part of the analysis. The univariate findings are provided in Table 30 below. Significant results are highlighted in gray.

Table 30: Univariate Results for Hypothesis H1.1

Item	**Case: IFIP**			**Case: Event**		
Univariate Results						
Which variables contribute to differences?	**F**	**Sig.**	**Observed Power**	**F**	**Sig.**	**Observed Power**
Comp	3.33	.040	.618	1.90	.154	.388
Problem	5.01	.008	.804	5.23	.007	.822
Cloze	14.3	.000	.996	5.51	.005	.843
Comp. Time	2.209	.111	.442	1.32	.272	.280
Prob. Time	1.15	.332	.241	5.21	.007	.821
Cloze Time	.503	.606	.131	1.75	.178	.361
Ease of Use	62.27	.000	1.00	73.37	.000	1.00

The results in Table 30 above indicate that the variables within which we consistently observe significant difference among treatment groups are the product measures for problem solving and Cloze tests and the process measure ease-of-use. Other variables in which we observe a difference in one case and not in the other case include the comprehension and the time taken to complete the problem solving task. These ANOVA tests indicate significant differences among treatment groups. The ANOVA procedures do not, unfortunately, indicate in which direction the differences lie.

Analyzing Differences Between Groups in MANOVA

The direction and significance of the differences between groups is the heart of all MANCOVA procedures. There are two methods for analyzing these differences; post hoc univariate tests and planned comparisons between sets of groups that are called "***contrasts***". Stevens (1992, p. 107) notes that in regards to the univariate approach:

> *"This (univariate) approach is appropriate in exploratory studies where the investigator first has to establish that an effect exists, However, in many instances there is more of an empirical and/or theoretical base... Here... the investigator has specific questions he wishes to ask the data. ... If we plan a small or moderate number of contrasts that we wish to test, then power can be improved considerably. While control on overall α can be maintained through the use of Bonferonni Inequality"*

The quote above indicates a tradeoff. The univariate tests provide a method for making all pair-wise comparisons. But because this procedure includes all pair-wise comparisons, we are accepting a larger α and a smaller β (power) than we would need to accept if we are not interested in all pairwise comparisons. If we know the comparisons that we want to make, we can then set up contrasts that minimize our exposure to type I and type II error simultaneously.

In light of this discussion, we set up a set of "simple" contrast for each MANOVA using SPSS. Note that "simple" in this case does not stand for "easy" but instead indicates a general type of contrast. Simple contrasts involve comparing the estimated means - which may differ slightly from the actual means shown in Table 30 - to a chosen "control" group. For Hypothesis H1.1 we chose the "Text" group as the "control" group for the contrast. Two Simple contrasts are created; One contrast comparing DFD & ER to Text, and one contrast comparing OOA to Text.

The results of these contrasts for each of the dependent variables over the two cases, are provided below in Table 31 and 32. Table 31 includes results for the product variables and Table 32 displays results for the variables related to effort. Significant contrasts are highlighted in gray and significance levels are calculated using the Bonferonni correction. Note that the difference ("DFD – Text" or "OOA – Text") represents the estimated difference between means. A negative difference, therefore, suggests that the Text group outperformed the other group (DFD/ERD or OOA) on that variable for that treatment group.

Table 31: Contrasts for Text vs. Graphic Comparison (Product Variables)

Contrast	Item	Case: IFIP			Case: EVENT		
		COMP	PROB	CLOZ	COMP	PROB	CLOZ
Text and DFD/ERD	DFD - Text	0.968	.793	-8.313	0.122	.619	-4.62
	Significance	.025	.360	.000	.368	.463	.007
Text and OOA	OOA - Text	0.160	2.573	-7.816	-0.561	2.039	-5.45
	Significance	.741	.003	.000	.149	.025	.003

Three effects were consistent across the 2 cases. As predicted, the Text group outperformed both the DFD/ERD and OOA groups on the Cloze test. This is natural as the Text group was exposed to the text description before taking the Cloze test. This result adds to the internal validity of the study by indicating that the Cloze test was not a simple exercise, and that individuals scored differently based on the information that was presented to them. In other words, the Cloze test was not simply a test of general knowledge, but a test that required some specific domain knowledge.

A second observation was that the DFD/ERD and OOA group outperformed the Text group in the problem solving questions, with the OOA significantly outperforming the Text group in both studies. The fact that DFD/ERD did not significantly outperform the Text group need not be surprising, since all diagramming methods are not equal. The difference between Text and DFD/ERD is simply more moderate and may been significant with a larger sample size. It seems natural to observe differences in how alternative graphical techniques score when compared to text. What is important in these findings is that participants provided with graphical descriptions (such as OOA) have been observed to possess the potential to significantly outperform participants provided with text descriptions on problem solving questions.

The third observation gathered from Table 31 was that no consistent difference was observed across the two cases in regards to the comprehension test. While it may seem strange to illuminate insignificant differences between methods, this finding may be the most important observation in the study. The lack of significant differences on the 12 question comprehension task, across two studies, suggests that the three presentation methods were roughly equal in their ability to transmit the "content" of the domain to the participant viewing the diagram or description. The claim that the information embedded in one or more of the presentation methods gave an advantage to one or more of the groups is not substantiated by the findings. What is interesting about this observation, is that while differences were not consistently apparent in comprehension tasks, differences ***were*** found in problem solving and Cloze tasks. This suggests that previous empirical studies focusing only on comprehension questions to measure "understanding", may have been missing important factors. The observation also suggests that the current instrument, with three dependent measures, is sensitive enough to pick up at least some of these differences. Hypothesis P3.1, described later in this chapter, will be an interesting test of the sensitivity of the instrument.

Having addressed the major findings with regards to the product measures, we can now move to address the measures of effort. Table 32 presents the results of the comparisons between the text "control" groups and the other two test groups. Significant contrasts are highlighted in gray and significance levels are calculated using the Bonferroni correction.

Table 32: Contrasts for Text vs. Diagram Comparison on Variables related to Effort

Contrast	Item	Case: IFIP				Case: EVENT			
		COMP Time	PROB Time	CLOZ Time	EASE	COMP Time	PROB Time	CLOZ Time	EASE
Text and DFD/ERD	DFD - Text	-3.21	.04	.718	-2.50	-.539	-3.38	2.18	-2.53
	Significance	.067	.981	.625	.000	.831	.058	.072	.000
Text and OOA	OOA - Text	-3.31	2.54	1.55	-3.60	-3.96	-3.04	1.72	-.370
	Significance	.072	.214	.319	.000	.142	.068	.180	.000

Two observations regarding Table 32 are interesting. First, the Text group perceived the text description significantly easier to use than the two graphical methods. This is contrary to our previous argument in Chapter 4 and our prediction in Table 31. While no reasons for this significant reversal are immediately available from the data, one potential explanation can be found. The fact that participants spend most of their early years learning to read and write text descriptions and very little (if any) time learning how to build and interpret graphical descriptions of domains, it seems natural that participants would find text descriptions easy to use. Another observation is that the group provided with the text description spent less time (not significantly less but consistently less) on the Cloze test than the groups given graphical models. Since the Cloze test was the last test in the study, there may have been a recency effect associated with the ease of use scale.

The second observation from Table 32 is the fact that there were no significant differences in the time spent on the three study tasks (comprehension, problem solving, and Cloze). Again, this "non-finding" may be the most important statistic in the table. The insignificant differences observed in the table suggest that the three treatment groups spent similar amounts of time on the problem solving task. If this is true, it is then possible to state that the improved scores associated with the graphical methods on the problem solving task are not directly developed due to significantly longer effort applied to the task. In other words, the higher scores in problem solving are not the result of taking more time. If we assume, as Mayer (1989) argues, that the level of understanding is directly related to problem solving scores, then we can suggest that graphical models are able to communicate a deeper level of understanding of a domain, in the same time as that provided by text descriptions.

Combining the observations above with those from Table 31 enables us to make the following statement:

The data collected in this study provides evidence for the conclusion that graphical models communicate at least the same amount of comprehensive knowledge, and lead observers to a potentially higher level of understanding regarding a domain, in the same amount of time as text descriptions. The data also suggest that the level of understanding gained by individuals viewing a graphical model, as opposed to text descriptions, varies with the graphical method being used.

The study, therefore, has provided evidence that supports Hypothesis H1.1. In chapter 4 section 4.1.2 several reasons were provided to explain these differences that were actually observed. These reasons included the improved ability of graphical models to communicate the structure between components in a system along with component behavior. Further, graphical models indicate the linkage between components directly and more clearly, and are less "verbose" than text descriptions. It is impossible from the data collected to indicate which, if any, of the reasons contributed to the differences that were observed. This is natural as the study was not directed at the *question "Why is there is a difference*" but instead on the question "*Is there a difference*?". This question will be addressed later in the discussion.

Having identified some differences between graphical methods and text descriptions, we will now turn our attention towards the direct comparison of the two graphical methods, namely "structured analysis" using a combined DFD and ER diagram, and the "object" approach using an OOA diagram.

7.4 Hypothesis 2: Structured Analysis vs. Object Oriented Analysis

Chapter 3 provided an outline of the differences between the more traditional "structured analysis" approach that combines Data flow and entity relationship diagrams and the "object-oriented" approach to systems analysis. Section 4.1 then argued that object-oriented approach holds a theoretical advantage over "structured analysis" for several reasons. First, the object-oriented approach (OOA) is able to represent both structure and behavior on the same diagram with a single grammar, whereas the Data Flow diagram and Entity Relationship diagram (DFD/ERD) require two diagrams and two grammars. The two diagrams are not obviously connected, and this difficulty may place individuals at a disadvantage, relative to an OOA diagram, when learning about a domain. A second argument suggests that human cognition naturally views the world and the things in it as "object-like" so that these thing possess both structure and behavior. Separating structure and function to describe a system, as is done in the structured approach, may make a lot of logical sense, but it may not be a "natural" way of viewing a system for many individuals. No attempt has been made in this thesis to separate the individual effects from these differences. This argument led to the following Hypothesis:

Hypothesis H1.2

H1.2: The OOA grammar will be better able to promote understanding in individuals interpreting a domain description than the combined DFD/ERD grammar

From this Hypothesis we predicted that:

1. **Comprehension scores for the DFD/ERD and OOA groups will be the same.** This prediction is made based on Mayer (1989). He suggests that groups given text descriptions will score similarly in comprehension tests to groups given conceptual diagrams. If this is true, then it follows that two groups, both with conceptual diagrams, are unlikely to generate large differences in comprehension scores
2. **Scores for the Problem Solving task will show significant difference with the group viewing the OOA diagram scoring higher than the groups given the DFD and ER**

diagrams. This prediction is made based on the argument made above that OOA will provide a superior conceptual understanding because it required the interpretation of only one diagram, and the "object" approach lends itself more naturally to human cognitive structures.

3. **Cloze scores for the group provided with OOA will be higher when compared with the group provided with DFD/ERD**. Given that neither the OOA nor the DFD/ERD group has seen the original text description, the Cloze test in this case acts as a test for the overall understanding of the domain. The argument that OOA promotes a higher level of understanding than the DFD/ERD grammar suggests that the Cloze scores should be higher for the OOA group.
4. **The time taken to complete the comprehension, problem solving, and Cloze test, should be shorter for the OOA when compared to the DFD group**. This prediction relies on the argument that the extra linkages required between the DFD and ER diagrams will require extra processing time. A further argument suggests that the "natural" OOA constructs will be understood more quickly, reducing the time necessary to complete the task.
5. **The group viewing OOA models will rate them as easier to use than the DFD/ERD** models for the reasons mentioned above.

To assess these predictions we again used MANCOVA analyses using the data from the intragrammar study. Fortunately, the MANCOVA analysis for these predictions is almost exactly the same as the analysis performed in the previous section comparing text and graphics. The only change we have made is in the contrasts we employ. The fact that we are able to address so many predictions with a single technique is a good indication of the power of multivariate techniques such as MANOVA for comparative analysis. We must accept, however, that the two sets of contrasts applied for this Hypothesis and the previous Hypothesis are independent and therefore the type I error terms are additive when viewing the study as a whole.

Much of the analysis performed in the previous section is still applicable for addressing hypothesis H1.2. For example, all results from assessments regarding the MANCOVA analysis in the previous section have not changed. The multivariate estimates produced by the MANCOVA and reported in Table 29 are also exactly the same. The univariate statistics provided in Table 30 are also unchanged as are the means for each dependent measure as reported on Table 28. Before moving into the new contrasts created for Hypothesis H1.2 The reader is invited to familiarize themselves with the information provided in Table 28, 29 and 30.

The new contrast applied to the MANCOVA analysis is focused on the OOA method. Two Simple" contrasts are created comparing OOA to DFD/ERD (DFD/ERD – OOA) and OOA to text (Text – OOA). The second contrast (Text – OOA) has already been analyzed in the previous section and is not reported here. This leaves a single contrast (DFD/ERD – OOA) for our analysis. Table 33 below displays the results of the contrast. Significant contrasts are highlighted in gray and significance levels are calculated using the Bonferroni correction.

Table 33: Contrasts for OOA vs. DFD/ERD (Product Variables)

Contrast	Item	Case: IFIP			Case: EVENT		
		COMP	PROB	CLOZ	COMP	PROB	CLOZ
OOA and DFD/ERD	DFD – OOA	0.883	-2.34	-0.686	.605	-2.65	1.258
	Significance	.021	.003	.649	.093	.002	.454

Three observations are important when considering Table 33 above. The first observation is that the problem solving score shows a significant difference between groups provided with OOA and groups provided with DFD and ER diagrams. Furthermore, this difference is significant in the predicted direction, that is, the OOA group scored significantly higher than the DFD group. These differences are of central importance as Mayer (1989) has indicated that higher problem solving scores indicate a more developed cognitive model and a higher level of understanding. The first observation from Table 33, therefore, confirms our original prediction of improved problem solving scores for the OOA group and suggests that on average the OOA group attained a higher level of understanding.

The second observation is that the mean score of the DFD/ERD group was higher than the mean score in the OOA group for the comprehension assess in both cases. In fact, in the IFIP case this difference was marginally significant. Mayer (1989) predicts only small differences in comprehension scores between methods. The reason for small predicted differences is that comprehension scores measure only a peripheral level of understanding. Comprehension questions can be answered directly from the diagram. The type of diagram, at this level of understanding, makes little difference. The fact that a participant can read from a diagram, does not necessarily imply that a participant understands the domain being represented in the diagram. Something more, beyond comprehension, is required.

It is also important to point out that while the difference in comprehension scores between groups in this study are small (less than 0.80 which is less than one question) the significant difference for the problem solving scores are large. What is interesting in the comprehension scores is that the OOA group scored significantly higher in the problem solving despite scoring lower - even significantly lower - than the DFD/ERD group in the comprehension test. This observation confirms the notion that the problem solving questions are measuring a different facet of understanding than the comprehension questions. It also enables the researcher to suggest that understanding may not be directly related to comprehension, as previous empirical researchers in system analysis may have assumed. Consequently, comprehension questions may not be a good indicator of the true understanding developed by participants interpreting a system analysis diagram.

The third observation was that the Cloze scores were not significantly different in the two groups, and the OOA group did not consistently score higher on the Cloze test. This result is contrary to the prediction made at the start of this section. These results are again somewhat surprising given the significant differences observed in the problem solving task. The lack of theory provides us with little ability to reason about the observed result.

Having discussed the three major observations for the product measures, we can now move on to a discussion of the differences between the variables related to effort for the OOA and DFD group. Table 34 below presents the findings for the "effort" variables. As usual, significant contrasts in Table 34 below are highlighted in gray and significance levels are calculated using the Bonferonni correction.

Table 34: Contrasts for OOA vs. DFD/ERD (Process Variables)

Contrast	Item	Case: IFIP				Case: EVENT			
		COMP Time	PROB Time	CLOZ Time	EASE	COMP Time	PROB Time	CLOZ Time	EASE
OOA and DFD/ERD	DFD – OOA	.241	-2.13	-1.10	1.10	1.52	-1.33	0.55	0.989
	Significance	.888	.253	.475	.003	.362	.455	.660	.001

The observation most apparent in Table 34 above is that participants in the DFD/ERD group rated the perceived ease of use of the combined DFD and ER method significantly higher than the OOA group rated the perceived ease of use of the OOA method. Our prediction at the start of this section suggested an effect in the opposite direction. This finding brings into question the hypothesis that objects are a “natural” way to view systems. If objects are so natural, why did participants find DFD and ER, which split structure and process, easier to use? This question is addressed to some degree, with the protocol analysis that follows this analysis.

Combining the observations made from Table 33 and 34 above enables us to make the following statement:

The data collected in this study provides evidence for the conclusion that participants viewing object oriented diagrams score significantly higher in problem solving scores, and not significantly different in Cloze or comprehension tests than participants viewing the combination of a Data Flow Diagram and a Entity Relationship Diagram. Further, the participants using the object oriented diagram do not take significantly different time to complete the study tasks when compared to participants provided with a combination of Data Flow Diagram and Entity Relationship Diagram.

Now that the differences between OOA and DFD/ERD have been observed, the next section turns to ask the questions of why these differences were observed.

7.5 Why are differences observed between OOA and DFD/ERD?

MANOVA is designed to establish whether differences between groups are present, but not to answer why these differences occur. To address the question of "why", another empirical method must be used. Earlier in Chapter 5, we outlined the empirical methods used to collect verbal protocol information of 12 participants. We will now turn to the results from this method to address the question of why the differences were observed.

Section 5.3 in Chapter 5, of this study outlined the methods used to collect the process tracing data. Verbal protocols were collected from 12 participants. Four participants used text descriptions to complete the comprehension, problem solving, and Cloze task associated with the EVENT case, four used the DFD/ERD combination, and the remaining four used OOA. Two of the four participants in each group were familiar with OOA, DFD, and ER diagram and two of the participants in each group were not. The protocols were collected and the section devoted to problem solving was transcribed. The focus on problem solving made it a natural choice to explore in depth as the problem solving task displayed significant differences between treatment groups in the MANOVA analysis. This section discusses the analysis of the protocol data. The protocol data was collected by the researcher and then transcribed. When the protocols were coded (as described later in this section) the researcher and an MBA teaching assistant worked independently and used printed transcripts for the coding.

The analysis begins with a predefined idea, on the part of the researcher, of developing a better understanding of why differences occurred between groups. In particular, why differences occurred between the OOA and DFD/ERD groups. The question of "why" did not emerge, therefore, as an important issue after considering the data, but instead was imposed on the protocol information from the outset of the analysis. It is important to note that the protocol information was being viewed in the light of the MANOVA analysis, and this view admittedly biases the researcher's perspective on the data.

Three items emerged from the detailed analysis of the protocols and related test documents. First, none of the participants provided with the DFD and ER combination made significant use of the ER diagram. The lack of use indicates a weakness in the DFD/ERD approach that required individuals to make connections between diagrams with different grammars. Second, the protocols associated with participants using OOA appear to have been more "structured" than the participants provided with DFD/ERD. This structure is evident in the more organized search and reference to the participant's memory of the diagram in the OOA protocols when compared to the DFD/ERD protocols. Protocols from several participants are provided to illustrate this point. The third item is not derived from the protocol, but instead is developed from study of the diagrams that participants created after completing the comprehension, problem solving, and Cloze tasks. A close look at the diagrams indicated that participants created diagrams closely resembling object-oriented diagrams regardless of which description (Text, DFD/ERD, or OOA) the participant received. The diagrams may provide supporting evidence for the claim that objects-oriented methods are "natural" cognitive building blocks. These observations are treated separately below.

Did Participants use the ER Diagram?

The first observation noted above was that participants provided with the DFD and ER combination were not using the ER diagram. As initial supporting evidence, all four participants who used the DFD/ERD combination made at least a passing reference to the DFD, while none of the four directly indicated the use of the ERD. The following protocol, while blatant, was typical of the DFD /ER user:

> **Researcher:** One thing, I did notice, before we move on, is that you didn't really look at the second diagram (entity relationship diagram)
>
> **Participant 5 (DFD):** *Oh, I forgot all about it.*

> **Researcher:** But you seemed to be comfortable answering questions just from the first diagram (data flow diagram).
>
> **Participant 5 (DFD):** *Sure, Yah. Can I see the diagram again? (laughs)... Hmmm.... (reads ER diagram)... maybe I answered that question wrong.... hmmm ... No.... OK I'm ready to move on....*

Participant 8, for example, noted at the end of the study that the ERD was "*difficult to understand*", and "*did not relate the to the other diagram.*" Similarly, Participant 7 also noted at the end of the session that "*Oh, I didn't use that diagram at all.*" In contrast the DFD was used extensively in problem solving as shown in the protocol made by participant 6 while reasoning about a problem solving question. Note the specific reference to items on the DFD

> **Participant 6 (DFD):** *... I can't think of other possibilities from these so I want to think at a higher level, whether there is any other possibility like seating plan, deployment sheet... not likely.... and employees...* ***I am thinking of that diagram, that DFD diagram, that's quite clear to me ...*** *(moves hands in air) here is operations and here, human resources....*

Only a sample from three protocols are provided here, but an analysis of the protocols for DFD/ERD participants showed an obvious "leaning" towards the DFD and a lack of concern for the ER diagram. Participants were observed to frequently reference the DFD during the comprehension task, and later to make reference almost exclusively to the DFD during the problem solving task.

The protocols provided a clear indication of the DFD dominance in the thought process of the participants. As further evidence, a somewhat more "objective" measure of the dominance of the DFD in the protocols was created. The measure simply counted the number of "specific references" in the problem solving protocol that were made to either the DFD or ER diagram. Since the two diagrams shared some conceptual items, the coding of these references was not straightforward. For this reason, a short description of the coding categories used to create the references is provided below.

The numbers of "references" relating to the different diagrams were collected across all five problem solving questions. A "specific reference" to the DFD was defined as any mention of the DFD itself or the processes and external entities provided exclusively in the DFD. These processes and entities include "scheduling", "creating deployment sheet", "developing staff requirements", "processing a promoter's application", "entering availability", the "operations department", "marketing department", "human resources department", and the "promoter" entity. The "specific" references to the ERD included any mention of the ERD itself, references to any of the attributes included in the ERD, or references to the cardinality of relationships between entities. References to items found in both the DFD and ERD were also counted in a separate category. These items included entities such as "program of events", "employee", "schedule", "events", "seating plan", "staffing requirements", and "availability." As a final measure, the items mentioned in the protocol but not contained in either of the diagrams were also collected and counted. These include items such as "ticket sales", "overtime", "low morale", "computer systems", "seniority", "reputation" and other terms.

Some examples of these specific references are provided below in Figure 12. The references are provided directly out of the protocols from individuals who were given the DFD/ERD. The words that indicate the specific reference have been highlighted in bold.

Figure 12: Examples of References to DFD and ERD

DFD references (each statement counts as one reference):

Participant 8: ... *first I look up at the top of the* ***diagram*** *for the* ***marketing department*** *and then .....*

Participant 7: ... *well the way that the* ***flow chart*** *looked, they did act independently... their link was through the* ***operations department....***

Participant 6: *Hmmm... (pause)... the* ***operations department*** *is not responsible for the schedule of events. It is responsible for the* ***staffing.***

Participant 5: ... *I wasn't thinking of the stuff that came before... All those other details... issues such as* ***scheduling*** *of the those people ....*

ERD References: (each statement counts as one reference):

Participant 8: ... *the* ***start time*** *of the event may interfere with ...*

Participant 7: ... *he didn't have the* ***skill*** *set required... just because he's available doesn't make him or her suitable...*

Participant 6: .. *one reason is that the* ***skills*** *are not matched for this event.*

Participant 5: ... *there might be a* ***time*** *conflict with* ***events****.. like very short* ***time*** *periods between events ...*

DFD and ERD Joint References (each statement counts as one reference):

Participant 8: .. *you may not have the* ***employees available*** *to staff.*

Participant 7: *I just think of this in terms of* ***staffing requirements****... so no staff... to accommodate additional capacity ...*

Participant 6:... *if the* ***seating plan*** *is incorrect* ***staffing requirements*** *will also be underestimated...*

Participant 5:... *Well the* ***staffing requirements*** *didn't need him.*

The coding of references using the method described above is admittedly subjective, so the absolute number of references in each category should be viewed with an appropriate level of skepticism. A second independent coder was used to provide some validation of these results. The results from the independent coder are provided in brackets in Table 35 below). The relative number of references in each category, however, does provide a useful indication of references most commonly used to answer the problem solving questions. The large differences between DFD and ERD observed below in Table 35, leave little doubt that participants were making more frequent use of the DFD than the ERD. The results of these references are provided below in Table 35.

Table 35: References to DFD and ERD

Participant	Number of References to DFD	Number of Ambiguous References	Number of References to ERD
5 (DFD/ERD)	8 (8)	4 (2)	1 (0)
6 (DFD/ERD)	22 (18)	8 (6)	4 (2)
7 (DFD/ERD)	12 (13)	3 (2)	0 (0)
8 (DFD/ERD)	16 (14)	7 (5)	2 (2)

(Results from independent coder provided in brackets)

The observation of protocols along with the counting of references to the DFD and ERD provides evidence that individuals were focused almost exclusively on the DFD and did not make a strong connection between the DFD and ERD. Three reasons for this observation can be suggested. To begin with, instructions on how to connect the two diagrams were not specifically provided to participants. Perhaps, if the participants were given more information on how to make these connections, then the ERD may have been used more. While this statement is undoubtedly true, it requires that we also accept the statement that the connections between the DFD and ER are not “obvious” or “natural” for participants. Accepting this statement implies that the OOA may be a more “natural” method than the DFD/ERD combination, which lends credence to the original intent of the analysis. The statement also ignores the fact that two of the participants (Participant 6 and 8) had previous experience with both DFD and ERD and still made no obvious connection between the diagrams.

A second reason for the observed differences might be that the problem solving questions did not require the participant to understand the information on the ERD. This statement would have more force if the DFD participants scored as well as the OOA participants in the problem solving. But the DFD participants scored lower. It seems plausible to suggest that the observed difference may be a result of the lack of structure rather than a bias towards questions focused exclusively on the DFD.

The third reason for the lack of use of the ERD may be that the ERD is more difficult to understand for novices than the DFD. The DFD is, therefore, more appealing for novice participants and consequently used more than the ERD. Some evidence for this view is provided in Table 39, where participants with more experience with ERD (Participants 6 and 8) were observed to refer more often to the ERD than participants without experience. Evidence for the preference of novice users for process models (DFD) over data models (ERD) has been provided by Vessey & Conger (1994, p.111) who state:

> *All the results presented here suggest that novice analysts found the process methodology easier to use than the data methodology, which in turn, was easier to use than the object methodology...*

The results observed this study corroborate those reported by Vessey & Conger (1994) in two ways. First, the DFD seemed to be preferred to the ERD by the novices in the study of protocols. This has been reported above Second, the ease of use scores observed in the MANOVA analysis rated the group using the DFD/ERD combination significantly higher than the group using OOA. If the DFD group was making significant use of only the DFD, then we can also corroborate that the process model was easier to use than the object model. While we have not observed why DFD is preferred to the ERD for novice users, we have supplied evidence that the DFD's were used more extensively than the ERD, and that participants developed no obvious connections between the DFD and ERD.

This observation can be used to create a plausible explanation for two observations in the MANOVA analysis. If participants provided with the DFD/ERD combination focused almost exclusively on the DFD, and did not take the time and cognitive effort to make the connections between the DFD and the ER diagram, then these participants would have taken less time than originally expected to complete the study. The observed times for DFD/ERD participants were shorter than expected as a direct result of the disregard for the connections made between the DFD and ERD.

Moving to the second observation, the focus on DFD suggests that the DFD/ERD participants lacked a structural view of the domain. If this were true then these DFD/ERD participants would tend to score lower in performance measures than the OOA group who had process and structure embedded in a diagram. If this were true then the differences in scores would be most evident in the problem solving performance where a conceptual view of both structure and process is more important. The faster times and lower problem solving scores observed in the MANCOVA analysis fit this description. One possible explanation, therefore, for "why" the OOA group scored significantly higher than DFD in problem solving is the relatively small use of the ERD diagram and the inability of participants to make the connection between the DFD and ERD.

Do OOA Participants Develop a More Structured Cognitive Model?

The second observation that became evident on examination of the protocol data was that the participants who were provided with the OOA diagram took a more organized approach towards answering the questions. It could be argued that the presence of a more organized approach to the answer of a question indicated the presence of well structured cognitive model that indicates a higher level of understanding. If it could be demonstrated that the OOA participants were better able to structure their analysis of the problem than the DFD/ERD group, it might be fair to conclude that the OOA enabled the development of a more sophisticated cognitive model than the DFD/ERD combination. In other words, the OOA may have facilitated a higher level of understanding in participants than the DFD/ERD combination.

The initial impression that the OOA groups provided more structured answers than the DFD/ERD groups was arrived at by simply reading through the protocols a number of times. The difficulty in addressing the issue of structure is how to develop the initial "impression" into a set of persuasive observations. Two approaches are taken in this section. The first approach is to attempt to quantify the "impression" using a variety of measures. These measures provide some insight into the differences between the two groups, but they are by no means conclusive and can be easily dismissed as small sample effects. A second approach is to provide the readers with a selection of excerpts from the protocols of participants in both groups. These excerpts are revealing, but may not be convincing as the same small sample effects may still be claimed. When the two approaches are combined, however, so that the protocols are backed by a set of more "objective" measures, a pattern is suggested and an argument can be formed. The reader is reminded that what the researcher is searching for in this section is not "proof" but rather "insight' into an explanation for possible differences between the DFD/ERD and OOA groups in the MANOVA analysis.

Eight empirical measures were created to attempt to quantify the impression of structure. The measures are discussed below. The first measure, problem solving score, is familiar since it was used previously in the MANOVA analysis. For this measure, we summed the number of acceptable answers to the five problem solving questions. These scores are provided in column 1 of table 36 below.

The next three measures simply recorded the average time taken by each participant to generate the first, second, and third answer to the five problem solving questions. To accomplish this, each of the protocols was broken into segments of 5 seconds. The segment in which an acceptable answer was started was recorded. For example, after breaking apart the protocol for participant 6 into five second segments it was found that the first acceptable answer to problem solving question #1 was started in the 5^{th} segment. This suggests that the first answer started after approximately 25 seconds (5 segments of 5 seconds duration). Since there were five questions, we could create an average time taken for the first answer for each participant. This average is reported in column 2 in Table 36 below. The averages for the second and third answers (if provided) are reported in column 3 and 4 respectively.

The next measure built on the analysis of "specific references" developed in the previous section. The data from Table 35 above was summarized into two measures: the number of specific references related to the diagrams (DFD, ERD or both) and the number of references made to items not included in the diagrams. These measures were then extended to the OOA groups. Since the OOA included only a single diagram, these references were more straightforward to code than the DFD/ERD references. The references to items not provided the diagrams included "vacations", "fire regulations", "food and drink", "employee sickness", "the grape vine", "licensing", "40 hours per week", "reserve force", and others similar to the references noted for the DFD/ERD participants in the section above. Again an independent rater was used to provide some validation for the coding of references. The results of this analysis are displayed in column 5 and 6 in Table 36 below.

Another measure included in Table 36 is number of "pauses" that were contained in participant's protocols. A "pause" was defined as either a short break in the monologue (silence), or the use of "stalling" words such as "ummm", "hmmm" , "aaah", "well... ", "OK", and "let's see". Of course, a considerable amount of subjectivity is accepted in developing this measure. The reason for collecting the "pauses" is to use them as a measure of the amount of searching (cognitive work) being done by the participants. The more stalling or pausing in the protocol, the less apparent the answers would seem to be. If the OOA groups was observed to pause less than the DFD group then this might provide some evidence that the cognitive model was more structured and "apparent' for the OOA group. Since some participants took longer to answer questions and had more time to pause, the absolute number of pauses is not a good indication of the relative amount of searching. To reduce the influence of time on the number of pauses, a ratio measure was created. The ratio created was the number of pauses divided by the number of 5 second time seconds used in answering the question. Column 7 in Table 36 below, therefore, reports the average number of pauses per 5 second segment over the five problem solving questions for each participant.

The final measure in Column 8 in table 36 is more subjective in nature. The measure is dichotomous (yes/no) and indicates if the protocol contains at least one instance where the participant directly indicated they were using the structure provided by the diagram to search for possible answers to the problem solving questions. If there was at least one instance then the measure was coded as "yes", if not then "no". Instances from each of the participants who provided direct indications of the use of the diagram's structure in developing their answers are provided below.

> Participant 6 (DFD/ERD*): ...I'm just thinking if there are other possibilities. Employees working schedule works on two variables – availability and skill. It's... I can't think of any other possibility from these so I want to think at a higher level, whether there is any possibility like seating plan, deployment sheet... not likely... and employee... I am thinking of that diagram, that DFD diagram... that's quite clear to me... (moves hands in air) her is the operations and here human resource department.. this is for skills this is for availability.... I can't think of any other reasons...*

> Participant 9 (OOA*): ... just trying to go around the parts,..around and you got the employees on the diagram, employees there, HR's there, operations down, then marketing up again. Employee is only connected to HR ... he's got skills, he's got availability. And then that's all connecting him with the HR part, and then operations is scheduling the number of people required. So ther's not much left....*

> Participant 10 (OOA*): .. what I am doing is I am trying to think about the main responsibilities of each department on the diagram... What the communication should be between the two and if that communication is lost, what the problem might be. So I kind of look at the marketing as developing the event which is going to be, say, the performers that are going to be on, say, advertising, say, the research in doing that... Whereas the HR department is more in charge of getting employees who have the right requirements to do the job ....*

> Participant 11 (OOA*): ... The seating... go down the other way... that picture... the promoter requesting the seating plan or depending the request of the seating plan to the operations department...Again... the issue, well the seating issue......*

> Participant 12 (OOA): *...Well I didn't see a processes that scheduled time before and after.. specifically... so maybe they are not ready... ready for this... There is nothing in the diagram for this... hmmm... maybe there is an impact on the marketing department... no they do not set the events up... hmmm.... I think that is it... I don't remember more from the diagram...*

The results of this analysis provided some of the strongest evidence that the participants in the OOA group structured their answers around the diagram more than participants in the DFD group. Surprisingly all four OOA protocols contained a specific reference to searching through the OOA whereas only one references was observed in the DFD/ERD protocols. The result of this analysis is provided in column 8 in Table 36 below. Particularly large differences between the DFD/ERD and OOA groups are highlighted in bold.

Table 36: Empirical Measures Derived from Protocol Data*

Participant	Previous Method Exper-ience?	**Column 1** Prob. Solv. Score	**Col. 2** Average Times to Create Answer 1rst	**Col. 3** 2nd	**Col. 4** 3rd	**Column 5** # refer to item IN diagram	**Column 6** # refer to item NOT IN diagram	**Column 7** Aver. pauses per 5 sec.	**Column 8** Searched diagram for structure
5	No	10 (10)	28	41.6	68	13 (10)	5 (4)	.84	No
6	Yes	12 (11)	74	111	135	34 (28)	3 (4)	.82	Yes
7	No	16 (14)	46	70	95	15 (15)	14 (11)	.70	No
8	Yes	16 (16)	47	72	98	25 (21)	6 (7)	.69	No
DFD/ERD Averages		**13.5**	**48.8**	72.8	99.1	**21.75**	**7.75**	.76	¼
9	Yes	22 (20)	44	75	117	47 (40)	14 (14)	.64	Yes
10	Yes	20 (19)	25	86	119	29 (25)	24 (20)	.83	Yes
11	No	14 (15)	28	70	100	38 (35)	5 (3)	.82	Yes
12	No	10 (10)	29	58.75	60	28 (26)	4 (5)	.88	Yes
OOA Averages		**16.5**	**31.7**	72.1	100.3	**35.5**	**11.75**	.79	**4/4**

*** Results from independent coder provided in brackets**

Several observations can be made regarding the results in Table 36. First, the OOA group scored higher on the problem solving task than the DFD group. This provides a small sample confirmation of the MANOVA results. The OOA groups were observed to create the first answer to the problem solving task approximately 17 seconds faster than the DFD/ERD. This advantage was not long lived, however, as the second and third answers were completed in similar times for both groups. The large differences in times to create the first answer provides evidence that the OOA group may have had an existing cognitive model that was more "organized" and more easily "accessed" than the DFD/ERD group.

In regards to the items referenced by the participants, the OOA participants were noticeably higher in referencing items that were contained in the diagram than the DFD/ERD group. The large difference suggests that the OOA group may have relied more heavily on the components in the diagram when formulating answers. As argued earlier, this observation may imply that the OOA participants applied more structure to their search for answers than the DFD/ERD participants. Further evidence that participants using OOA relied more heavily on the diagram than the DFD/ERD participants is provided in column 8. As noted earlier, all four OOA participants made specific reference to the OOA diagram whereas only one participant did so for the DFD.

One further finding is important to note. Column 6 in Table 36 above indicates that the OOA group were more successful in assimilating items that were not included in the diagrams into their answers than the DFD/ERD group. This is particularly true for those participants (Participant 9 and 10) who had previous experience with the OOA model. The assimilation of items not included in the diagrams may indicate a deeper level of that intertwines information from the diagram and personal experience.

The measures and results discussed above provide us with only a glimpse into the understanding of a participant. While the actual level of understanding is not directly observable, these measures can provide some insight into why differences between OOA and DFD/ERD were observed. The discussion above suggests it may be plausible to attribute some of the difference between DFD/ERD and OOA participants to the notion that the OOA diagram supplies a structure that is more readily applied to the way individuals think about a domain than the DFD/ERD combination. The OOA may be more readily applied because it encapsulates both behavior and structure of things in the domain, or because it breaks the domain into recognizable chunks - what are call "objects" – that more closely approximate the cognitive constructs individuals use to develop understanding. Whatever the nature of the advantage, it is clear that differences do exist in the number of references, the time to first answer, and the direct references to the diagram and structure.

Having provided some quantitative data on the differences between OOA and DFD/ERD participant's protocols, it is useful to consider the actual protocol data itself. A samples of six protocols are provided below for question number 2 in the problem solving task. The same question is used throughout all the protocols. The question is:

Suppose that it is a very busy two weeks for the entertainment complex with events scheduled for every day of the week. What problems might be encountered due to this busy schedule?

The entire protocols for the answer to the question above for all eight participants (4 DFD/ERD. 4 OOA) are provided on the four pages below. It is very important to note that none of the participants had the diagrams in front of them while answering this question. The participants were using only their own understanding and memory to create these answers. These example protocols are provided here because they provide the richest and most candid view of participants' understanding. Before each protocol is listed, a short description of "interesting" items in the protocols will be discussed. These items are highlighted in bold in the actual protocols to reduce the search time necessary for the reader.

The first protocol to note is from participant 9 (OOA) in Table 37. Note the structured approach taken by numbering the answers and the diligent search through the objects in the OOA one by one as they appear in the diagram. This can be contrasted by a somewhat less organized approach observed for Participant 6 (DFD/ERD). Pay particular attention to the question that the participant asks him(her)self, and the effort required to address this question. This "working out" of a solution was not as apparent in the protocols from OOA participants as it was in the DFD/ERD protocols.

The next two protocols that follow participant 11 and 6 are the protocols from participant 10 (OOA) and 7 (DFD/ERD) as provided in Table 38. Note again, the structure that the OOA participant brings to the discussion by categorizing the effects. Participant 10 makes less use of the OOA than Participant 9, but the organization is evident. This is contrasted with the approach by Participant 7 (DFD/ERD). Note that Participant 7 makes very little use of the DFD and relies more on external items such as labor laws, overtime, exhaustion, double-time, and other factors not included on the diagram. No evidence of a systematic search through the DFD is present in participant 7's protocol.

Table 37: Protocols from Participants 6 and 9 for Question 2

<table>
<tr><th>Participant 9: OOA</th><th>Participant 6: DFD/ERD</th></tr>
<tr><td>

*Good, I can come up with quite a few on this one..... **One, marketing department reject promoter's request for event due to lack of space**.. that's the only thing. Then he requests a seating plan and is extremely crowded and a seating plan is suggested by the promoter is not appropriate....*

*The operations department also makes an objection. **Two, operations department does not approve promoter's highly dense seating plan** for reasons of either employee deployment or security or safety....*

Operations department now has a problem, cant compact the seating arrangement suggested by the promoter. Goes to the operations department . . . yes, no, processes it, clears the deployment schedule, now the high turnover of customers, too many employees, employee availability could be an issue, because too many people are coming, like a vacation time, therefore employees must be also off on vacation, some of them, so I'd assume.

***Three, employee availability is affected by the busy week**. HR department may not be able to support the requirement for employee deployment. Employees, from the HR. One , two, three, four, ... employees are on vacation. (reads question)...employee, HR, marketing, operations.. promoter ....*

Researcher: What are you thinking about?

*I am trying to go around.(laughing)... Umm ...it's like these things are the points that hit me straight away. **I reckon because it was clearly defined on that map**. Marketing is rejects a clear difference, will approve or disapprove. So those things straight away came to the head. **But now once those clear differences are gone, now I have to go deeper inside to see what kind other potential problems could exist and um ...and now without the map I cannot think what was on it**.*

Um.. employees' schedule was there, employees... and would conflict come in there too in the sense that employees could argue over their vacation time. Since not everyone is being forced to work ...

Deployment ... OK employee gets aggravated and everybody in the department gets slowed a little bit. The efficiency comes down. All departments get overload and may lead to conflicts. Which may be ... I don't remember it on the diagram... some side issue. Let's see what else. It's a sports complex... events are schedules for every day of the week .. that doesn't make a difference .. (reads questions again) ... (pause) .. No I think I am running out of ideas here...

</td><td>

... It is very busy (rereads question)... OK so, because it is so busy a schedule.. the schedule was.. was.. actually the schedule... whether it is busy or not it is responsible for the marketing department because the marketing department is responsible for which event will be held or not... Ahhh.. so ... Maybe marketing department accept too many events... but marketing department just approve or not approve events. They put events in the program events store... Who is responsible for the schedule?

Hmmm... (pause)... the operations department is not responsible for the schedule of events. It is responsible for the staffing. Humana resource department is not responsible for the schedule. Human resource are responsible for the employees. So it must t be the marketing department who did not schedule well for the events. ..

Oh yah, they check, marketing department check when a promoter submits their programs, their events, they must give the marketing department the vent name an d time. The marketing department may not look carefully about the time.... so that the cause so busy in the schedule.....

I think it is marketing department. Or maybe at that time a lot of promoter. Maybe it is a seasonal things, I don't know, a lot of promoter submit their applications for entertainment. Maybe, suddenly a lot of promoters come. ... That's it ...

Researcher: You might be misinterpreting the question slightly. Could you read the question one more time. If you could? Just read the question aloud.

(rereads question) ... Ohhh, what problems might be encountered... Oh I see.. yah... What problems... OK. Ummm ... problems will be not enough staff because. Because.. the promoter will present seating plans and have staff requirements and those staff requirements will require staff to work on that. But if the staff is limited – staff depends on their availability and their skills – so if too many events happen, staff may not be available for certain events...

or... available staff will not have the skills required for the event ... Ok that's it. I'm finished.

</td></tr>
</table>

Table 38: Protocols from Participants 7 and 10 for Question 2

Participant 10: OOA	Participant 7: DFD/ERD
Um, well.. um... from the... ***I could categorize this. First the employee standpoint.*** *That view. I would look at things about availability of employees, and overtime (writing)... things like I was mentioning in the past one where an employee had already worked their 40 hours so they weren't on the priority list anymore. Um ... other things along those lines, other events, at least the ones I have worked on, tend to be pretty unpredictable what's going to happen. You're going to have a reserve force and then you don't know how late the event was going to go, and ah ... the amount of unpredictability or variance , I would call it, on the day of the event.* ***The next set of problems not employee related would be more facilities related*** *I'd say. let's say we are down at GM Place and we have a hockey game on one night, a basketball game the next night, a car show the next night or whatever. Ah.... so let's say the hockey game goes into triple overtime and you have to have your crew down there taking the boards down for the basketball game. Umm .. those sorts of facility type issues. I would come up with next....* *The next I would look at other resource related constraints .. resource and supply constraints. I've looked at employees as a resource before and I have a bank of employees and we can yank them out whenever. But, there are a whole lot of things you need besides employees. You've got to have your umm.. well your food for instance... you gotta have your concessions. Ah... assuming you are having them for events...* *Sports event... let's see what do you need to have for a sports event. let's see you have your giveaways and prizes, your whatever, all of these unknown things, umm ... unknown resources. I'm going to roughly categorize them roughly under there. ...* *The next thing I would think of is ah.. what if something goes drastically wrong one night. Ahh... what is power goes out in the complex or something. It kills the event the next week or the next night Or your computer system goes down, whatever...*	*... Well ...having... having adequate staff... adequate number of staff... available... adequate staff with skills required ....* ***Researcher****: What are you thinking about?* *Well... I am just trying to imagine what... the ummm.... possibilities might be.... So if I understand this correctly... you are very busy... events everyday of the week... so for me so far it just looks.. it ahh... numbers... and skills ... what else could be ummm... a challenge...* *Umm... observing... ummm.* ***labor laws*** *.. so you are not working people more hours per day...(pause) and... staff cannot work in excess of ... I don't speak clearly when talking aloud (laughs)* *... you don't want your staff working 16 hours... I mean you can always have them but that's not... ummm. the way to go... well this is an entertainment event... and you want your staff to be... have the right attitude for work...and they won't have the right attitude if they are overworked. They won't ahhh... be terribly receptive to serving the public..* ***balancing work hours and exhaustion****.. staff work hours and exhaustion.. to provide.... and I think that's about all.* *So paying for people, paying overtime, budgetary , and if you don't have enough people... and even if you do.. a week like that you could really fork over a lot for* ***overtime. It could really hit you hard. Especially if you've got a .... if you are going into double-time....*** *I'll leave it there.*

The protocol from Participant 11 below in Table 39 is interesting because the participant had never used an object-oriented diagram. Even with the lack of training, note how Participant 11 searches through the diagram to look for answers. The organized search provides some indication that Participant 11 is relying on the diagram to generate answers. This type of search is again not evident in the corresponding protocol from Participant 8 (DFD/ERD).

Table 39: Protocols from Participants 8 and 11 for Question 2

Participant 11 (OOA)	Participant 8 (DFD/ERD):
Doesn't have to do with marketing anymore, unless they are trying to market the events...Well I guess one of the problems could be that the they could be accidental conflicts scheduled by the marketing department, or potential acceptances of proposals that.. or confusion ***The seating.. go down the other way... that picture.. the promoter requesting the seating plan or sending the request to the seating plan to the operations department...*** *Again... the issue.. well the seating issue, is that even if they are all booked up, they may not have enough seating... depends if somebody has actually planned that out... because if there is a whole bunch of events going on at the same time, not., well, yah.. at the same time... or at a sports complex... that is based on sport complex being a big complex with different rooms, not just one big room...* *Ummm, what else could happen... ummm.... let's see... they can be understaffed, unless they are willing to pay overtime... depending on how many people they have... ummm. they might be understaffed in the sense that there are staff or certain staff.. don't have the skills necessary for all .. dealing with all the events... Umm.. there could be scheduling problems.. Going back to operations, if there is an error in determining all of the things that need to be done... Umm... the operations department really has to focus on or get all these projects under their belt and figure out how many people they will need and .... where the.. where they are going to need them or when they are going to need them.? .* *And if there is a delay between the units such as the operations department and human resurfaces.. like.. operations know how many people they need... but they don't get the information to human resources early enough, ummm. because human resources created the schedule ...then they won't have a staff to take care of the problems.*	*OK...ummm... so there are a lot of events in the period... and with so many events.... it might be the case that you wouldn't be able to get them all in.. you know...you might have to turn down or reject some of the promoter's requests... that would decrease revenue... what else... OK.. ummm....* *... there's also the problem of employees... how many employees will you need to cover all of these events... you might not... not have enough and create a shortage with all the events happening.. happening too close....* *I guess there's also the possibility... that the employees are working... working too much so that they are ..well... going over the maximum.. the maximum amount they can work...* *Ok...umm.. there are other things like it may be difficult to produce the... schedule.. you'd have to use a computer system for a big event... so there is the possibility of scheduling conflicts if you weren't able to keep track...* *... and I guess that's it. That's all I can come up with...*

The final two protocols are from participant 12 (OOA) and 5(DFD) as shown in Table 40 below. These protocols are relatively short and provide little opportunity for insight. They are provided here in an effort to show a complete view of the answers to a single question. Perhaps the most interesting part of these protocols is provided by Participant 5 (DFD/ERD) who notes some confusion developing regarding his(her) answer. The "messy' comment at the end of the protocol indicates Participant 5 may not be searching systematically through the diagram.

Table 40: Protocols from Participants 5 and 12 for Question 2

Participant 12: OOA	Participant 5: DFD/ERD
OK... Not enough employees... this is two weeks...umm...every day of the week.. so... marketing shouldn't have any problems... they don't have too much of a workload.... (long pause)... ***Researcher:** What are you thinking?* *... operations do the seating plan... yah.. I don't know how efficient they are... in that.. they may not be able to do the seating plan.. or the scheduling... ummm* *... the employees get two weeks in advance notice... maybe they should get more... it depends on how many employees there are .... that's it...* ***Researcher:** What are you thinking?* *... I think that's all I can think of.*	*ummm... there might be time conflict with events.. like very short periods of time between events which would make .. it difficult for employees to work or whatever ...(long pause)* **Researcher**: What are you thinking about? *ummm... there may not be enough people to cause not everyone is available everyday and ...* *.. schedule for every day of the week... problems encountered due to this busy schedule.... ohh... they may have to pay people over time...ummm ... there may not have enough people to cause not everyone is available everyday and ...* **Researcher**: What are you thinking about? *I'm just reading the question over .. (rereads question)* ***... it gets really messy ...(laughs) I don't know.. that's it...***

In my opinion, the DFD/ERDD protocols lack the cognitive structure associated with the protocols observed from the OOA participants. It is freely admitted that no definitive evidence of a difference in the structure of the cognitive models developed by participants in either the DFD/ERD or OOA has been supplied in this section. What has hopefully been accomplished by listing these protocols and the related quantitative measures is to highlight some differences in the way the participants organized their answers. The observation of this researcher that the protocols from the OOA participants showed more structure than any of the three protocols from the DFD/ERD method suggests a causality between the use of the OOA method and increased structure of answers. The increased structure may be the advantage that the OOA provides over the DFD/ERD combination in order to deliver the significant differences in problem solving scores as reported earlier in this chapter. The next section provides some evidence for why the OOA may possess an advantage over the DFD/ERD method.

Are Objects "Natural" ?

The third observation that I have chosen to focus on is the similarity between the representations developed by all participants. The representations were developed in response to the following request that was read aloud to each participant:

> **Assume that a person will be coming into this room and that your job is to explain to this person the case that you have just learned about. You can assume the new person has no knowledge of the case and no knowledge of any formal system analysis methods or technique. In preparation for this explanation, I would like you to create a diagram that will help you explain the system to the new person.**

There were no constraints placed on the development of the diagram. The researcher stayed in the room to answer participant's questions. Three participants asked whether they "had to redraw the diagram" to which the answer was "No". Participant 4 asked if he(she) could use "text" and the answer was "Yes". In general, the participants found little trouble in creating the representation. This section reports on the results from the analysis of the representations. The representations created by the participants are collected in Appendix G.

Two observations were immediately obvious after viewing the representation created by the participants. Not surprisingly, the first observation was that the diagram created by a participant was generally similar to the diagram or description that was originally handed to them. For example, participant 4 was originally given text and then chose to create a description using only text (see figure 12 below). Participant 5 was originally given a DFD/ERD and then drew a diagram that had many of the essential features of the original DFD (without the processes). Participants 9, 10, and 11 were given the OOA diagram and all three created a diagram very similar to the original OOA. This observation suggests that the participants were guided by the representation they were handed. This is an important point. This finding reveals that diagrams are powerful mechanisms for structuring a person's thoughts about a system. Diagram not only communicate information, they help to organize it for the person viewing the model. Researchers should understand that system analysis models not only describe a system, they define how individuals think about the system and are therefore a powerful tool for communication.

Figure 13: Representation Created by Participant 4 (Text)

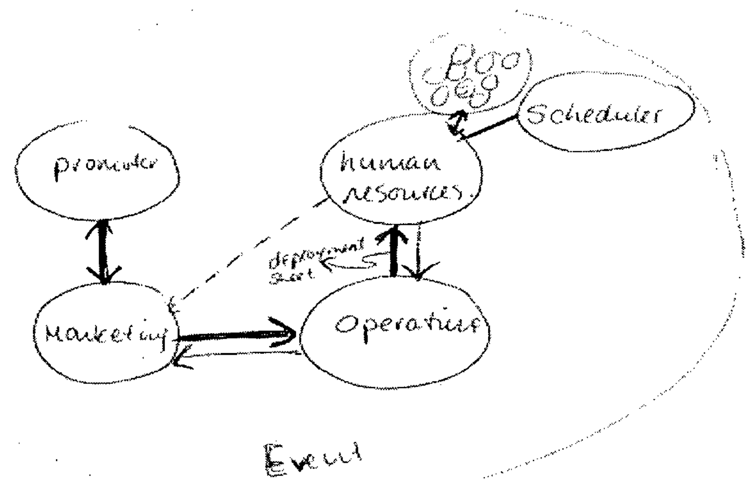
Participant 4: Text

1) Promotions Dept obtains/receives & application from Promoter
2) Promotions checks itinerary of events
3) Promotions accepts/rejects application
4) Promoter sends in seating plan to Ops Dept.
5) Ops reviews seating plan & develops idea of supporting personnel
6) Ops & Promoter review staffing requirements
7) Ops created deployment sheets & transferred to the HR Dept
8) Employee availability cards & transfer them to the HR Dept.
9) HR depts schedules employees using the employee availability & deployment.
10) HR dept makes sure no one works 40 +
11) HR prepares a final schedule

A closer look at the contents of the diagrams revealed a second observation regarding the nature of the diagrams created by participants. Many of the diagrams contained elements similar to “objects”. For example, the representation created by Participant 2 who was provided with the test description, contained the objects “promoter”, “marketing “, operations”, “human resources”, and “employees” in an arrangement that was very similar to the OOA used in the study. While the attributes and services were not contained in the diagram from Participant 2, the connections between objects was maintained and again mirrored the connection made on the OOA used in the study. It is important to note that participant 2 had no formal training in either DFD, ERD or OOA. The diagram for Participant 2 is provided below in Figure 13. Another important note is that Participant 2 did not perform well in the problem solving task scoring only 8 which was 2nd worst in the 12 participant group. It seems that Participant 2 was able to create an “object-like” diagram but did not benefit in the same way that participants who originally viewed OOA did. This suggests that interpreting from a model is different from representing the model.

Figure 14: Diagram Created by Participant 2 (Text)

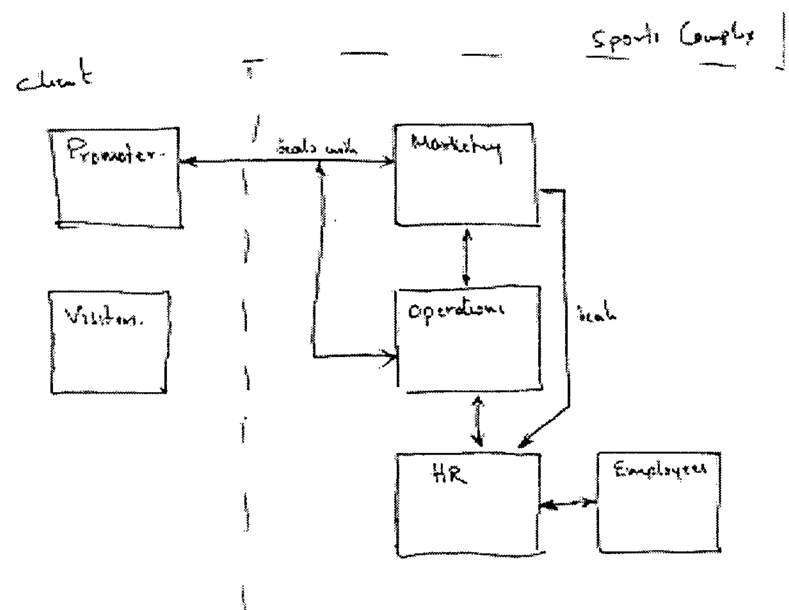
Participant 2: Text



The diagram created by Participant 2 (Text) above is very similar to a diagram created by Participant 9 (OOA) who had the benefit of OOA and previous knowledge of DFD, ERD and OOA. The diagram created by Participant 9 is provided below in Figure 14.

Figure 15: Diagram Created by Participant 9 (OOA)

Participant 9: OOA



Several other participants created diagrams that were object-oriented and included some attributes and services. Participants 10 (OOA) and 11(OOA), for example, developed diagrams similar to Participant 9 above, but with some added services. It should be noted that while

Participant 10 had the benefit of prior experience with OOA, Participant 11 did not. These diagrams are shown below in Figures 16 and 17.

Figure 16: Diagram Created by Participant 10 (OOA)

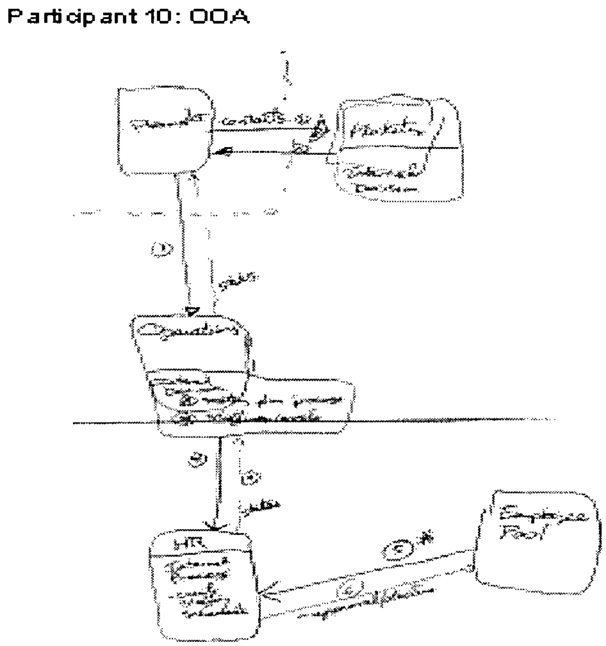
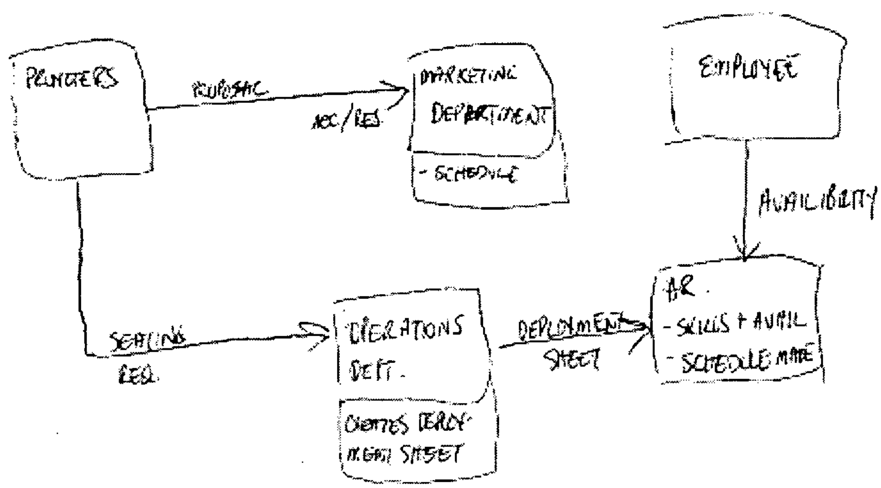
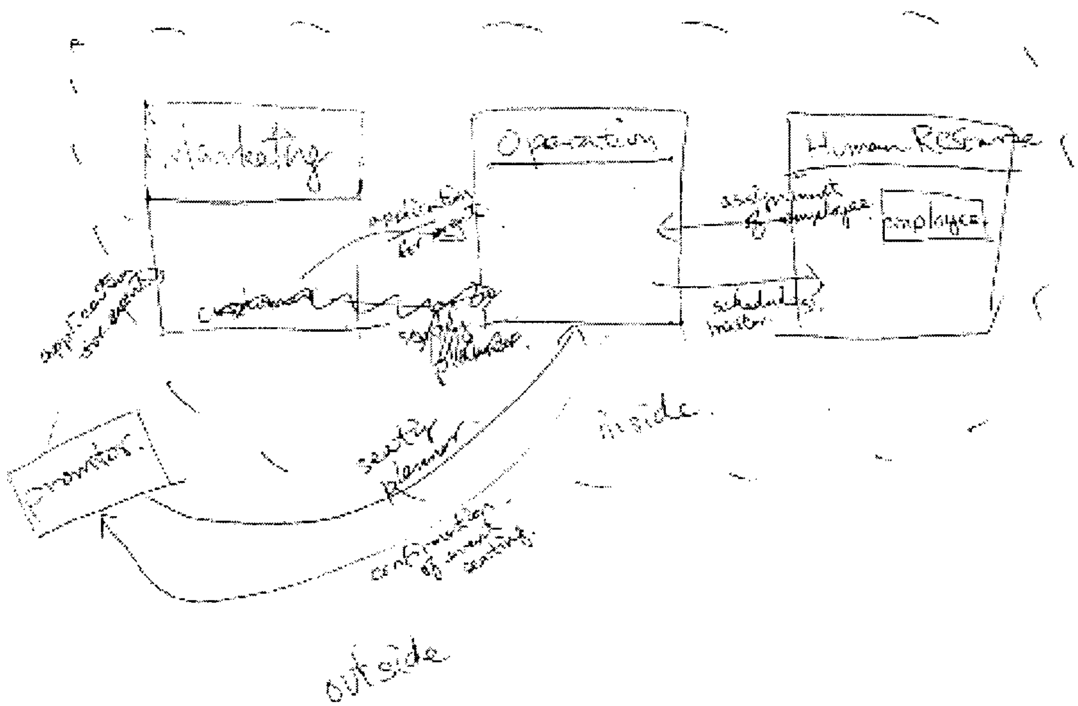


Figure 17: Diagram Created by Participant 11 (OOA)



Somewhat surprisingly, Participant 1 (Text) also produced an object-oriented diagram with services. The diagram is only somewhat surprising as Participant 1 had experience with DFD, ERD, and OOA methods. The diagram for this participant is provided below in Figure 18.

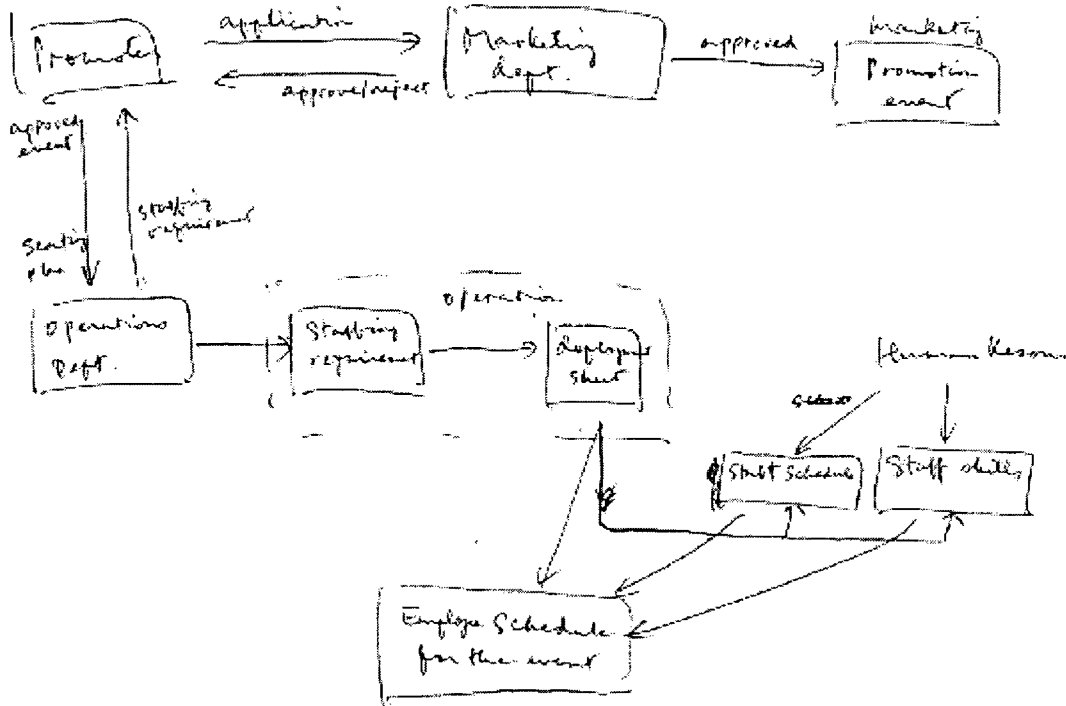
Figure 18: Diagram Created by Participant 1 (Text)



To this point in the analysis we have looked at 6 different representations, 5 of which have been object-oriented. Not all of the diagrams, had obvious object-oriented foundations. As an example, the diagrams created by participants originally using DFD/ERD more closely resembled DFD. No participants created diagrams resembling ERD, another indication that the ERD was not widely used. There were important differences, however, between the original DFD and the way that the participant's represented it.

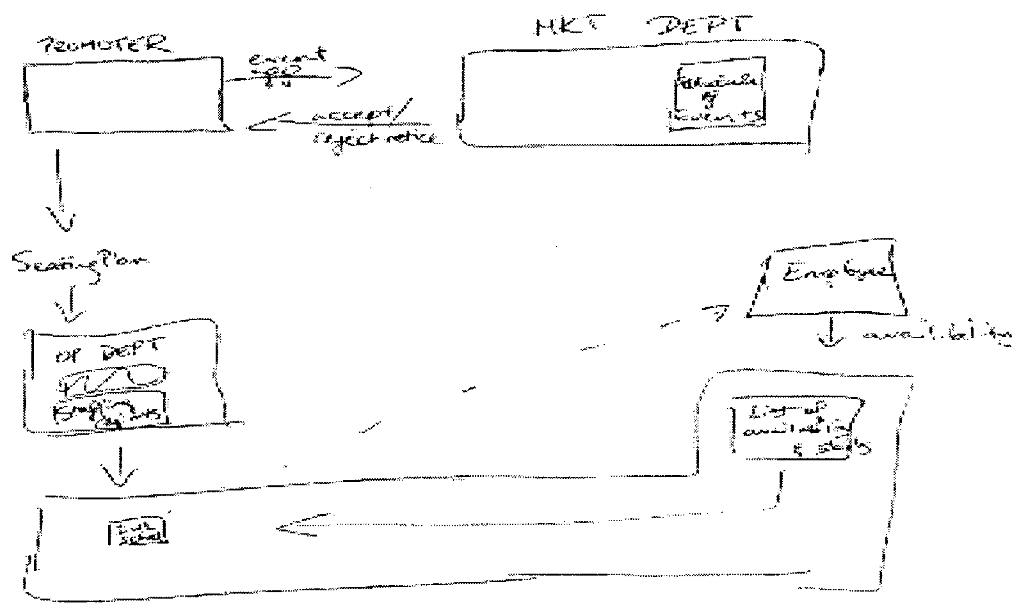
In general, participants originally provided with DFD did NOT connect processes with Data flow and did not define external entities. Instead, they labeled process boxes as "Marketing Department" or "Operations" and then connected the boxes with a combination of either data flows or in some cases names of processes. Little or no distinction between external entities (like "promoter") from internal structures (such as "Marketing Department"). An example of this is provided in the diagram created by Participant 8 (DFD/ERD). This diagram is shown below in Figure 18. What is of great interest in this diagram is the obvious confusion that the participant has in consistently applying the "simple" constructs associated with the DFD. Some data flows are names, some are not. Some boxes are departments, other are processes, and still others are external entities. Consider how a department is defined in the diagram. At one point a department (operations) is defined by dotted lines (as it is in the original DFD), and at another point the operations department is defined as a box. Continuing this investigation, the marketing department is again defined by a process box, while the human resources department is indicated simply by two directed arrows.

Figure 19: Diagram Created by Participant 8 (DFD/ERD)



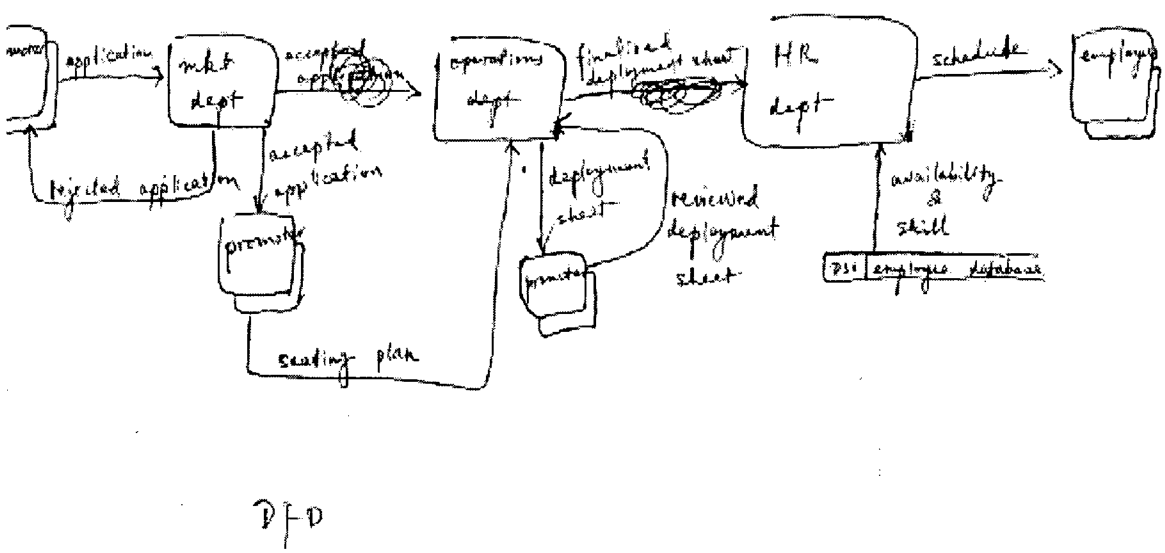
Participant 5 (DFD/ERD) shows similar confusion in how the constructs of the DFD are applied as shown below in Figure 20.

Figure 20: Diagram Created by Participant 5 (DFD/ERD)



Participant 3 (Text) had some experience with ERD and DFD and decided that he(she) wanted to create a DFD to aid the explanation. Note the "DFD" label at the bottom of the diagram. The resulting "DFD" is shown below in Figure 21. Note the confusion in applying the process box construct. Participant 3 is actually taking an object-oriented approach by naming the **departments** (objects with behavior and structure) rather than naming the **processes inside the departments** which is the basis of the DFD approach. The "process" boxes labeled "marketing department", "operations department", and "HR department" would normally be indicated on a DFD with dashed lines that surround the processes within the department. The encapsulation of the behavior of the department suggests into the single box labeled department in a construct more similar to objects than any construct in the DFD grammar.

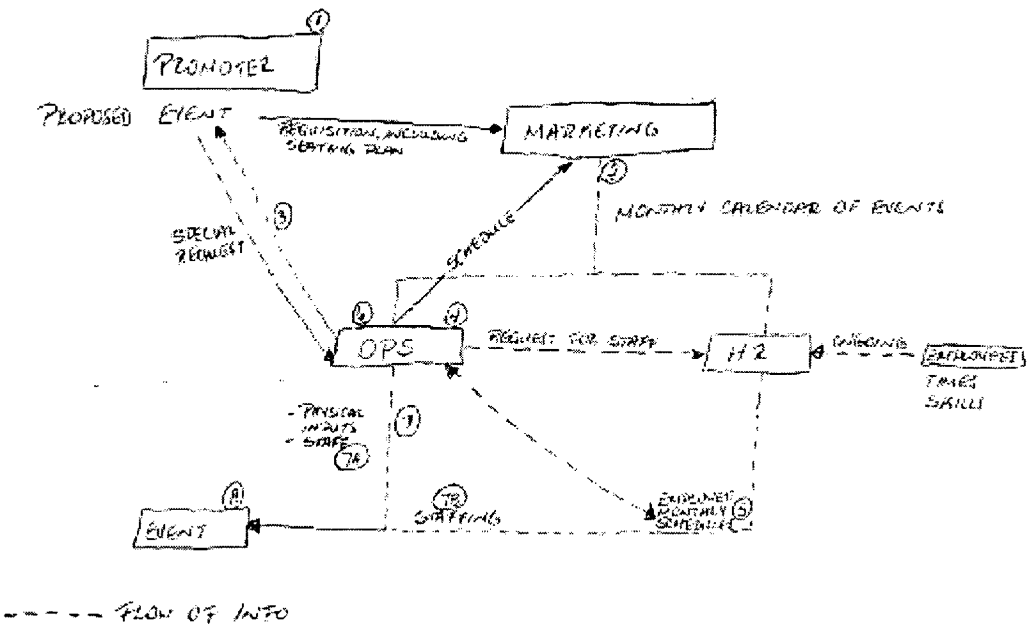
Figure 21: Diagram Created by Participant 3 (Text)



The three quasi-DFD diagrams found in Figures 19, 20, and 21 above provide evidence that the DFD grammar may be difficult for novices to apply, and that participants gravitated towards constructs that possessed more "object-like" features (primarily encapsulation). Even when participants started with a correctly specified DFD, the diagrams they created for explanation began to utilize object concepts. This observation provides strong evidence that the "object" method may be a more "natural" way for individuals to view the domain.

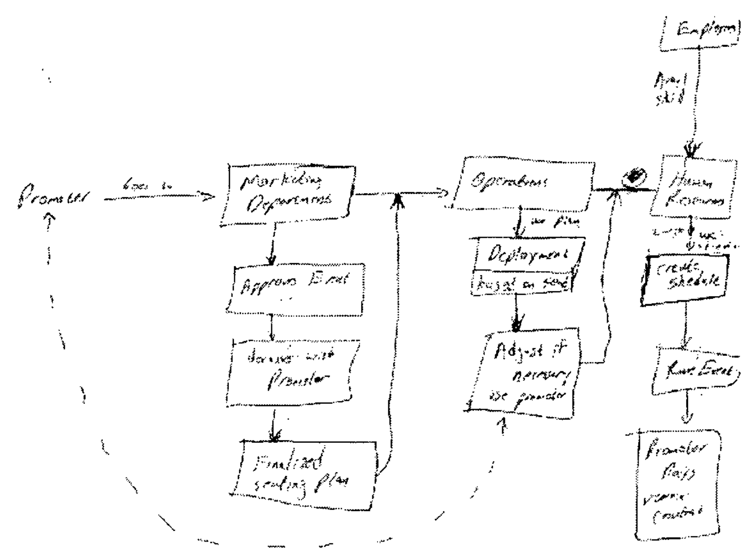
The observations made above can also be furthered by a look at the two remaining diagrams from participants 7 ((DFD/ERD) and 8 (DFD/ERD). These participants did not attempt to redraw the DFD but instead chose to represent the domain using their own techniques. Participant 7 again identified the main departments and then linked the department with flow of data. The sequence of the data flows was indicated by numbering the flow. This is shown in Figure 22 below.

Figure 22: Diagram Created by Participant 7 (DFD/ERD)



Participant 8 took a different approach. The starting point was again the departments, but in the diagram the processes within each department were specified. The sequence of processes within each department was outlined and then the flows between departments were defined. This is shown in Figure 23 below.

Figure 23: Diagram Created by Participant 8 (DFD/ERD)



Several things seem remarkable in the analysis of these diagrams. One observation is the degree to which the OOA and DFD participants differed. No two participants originally given the DFD diagram chose to use the same DFD constructs for their diagram, whereas three of the four OOA participants stayed with the OOA constructs. This observation lends support to the idea that participants were not using the DFD diagram for reasoning about the system. If they were, then why would the participants consistently represent the DFD incorrectly or change the constructs used by the DFD? Second, the congruence of the OOA diagrams suggests that participants were comfortable in using the object constructs to explain the domain. This lends support to the claim that objets are "natural" cognitive constructs, although this should be tempered with the realization that the DFD/ERD combination may have been lengthily to remember and, therefore, more difficult to represent. Third, the observed gravitation of the DFD/ERD participants towards "object-like constructs again provides an indication that object constructs may be a more "natural" cognitive structure than data flows and processes.

Summary of Results from the Protocol Analysis

The analysis of protocol results began with the goal of gaining insight into why differences between OOA and DFD/ERD participants were observed. Three observations from the protocol study were investigated. First, it was recognized that participants provided with a combination of the DFD and ERD focused almost exclusively on the DFD and did not make connections between the DFD and the ERD. The lack of knowledge regarding the structure of the data stores and data flows may have placed the DFD/ERD participants at a disadvantage in comparisons with the OOA participants. This observation suggests the system analysis researchers' expectations that individuals can easily connect concepts in separate diagrams that are based of different grammars may be optimistic for novice participants. In this study, participants displayed a higher level of understanding when showed a single diagram with a consistent grammar. These findings substantiate claims made by Coad & Yourdon (1991, p. 24-25) :

> *The data flow approach gives very weak emphasis to the data store. And the weakness is duly acknowledged by many authors describing data flow methods. So, many authors have tried to tie information modeling (ERD) into data flow diagramming, to compensate for the weakness. It's an academically pleasing idea (two perspectives, one system under consideration). Yet even in books, the connection is very weak.... And more important than what a textbook says, in practice the connections are virtually non-existent... Each analyst needs the benefit of both perspectives. In practice separate models keep critical issues too disjoint. And although CASE tools support could help somewhat, the analyst with data flow diagrams works primarily with a model that hides the impact of the data structure.*

The question is, if the DFD/ERD combination does not work well but is used widely, does the object model provide us with any hope that object oriented analysis will work any better?

Fortunately, observations in this section have provided justification for potential in the object model. For example, the observation made in this section regarding the relative “organization” of the protocols associated with OOA participants as compared to DFD/ERD participants. Several indicators of the organization of OOA participants were presented. These included faster times to the first answer for OOA participants, many more references to items in the OOA diagram for OOA participants, and more direct references to the OOA diagram while searching for answers. The researcher gained the impression that OOA participants used the OOA diagram as a “touchstone” for their search for answers, returning to the OOA when lines of inquiry wound up empty. The constant and consistent use of the OOA diagram while answering questions suggests that the object model might be viewed as a more “natural” cognitive structure than the DFD/ERD combination. In this way, the OOA participants were able to provide a more structured cognitive model using the OOA as a base than the DFD/ERD participants.

Evidence for the claim that objects are a more “natural” grammar than the DFD/ERD was also provided by the observations of diagrams created by the protocol participants. The OOA participants created diagrams consistent with the OOA grammar and two of four Text participants created “object-like” diagrams. The DFD/ERD participants, on the other hand did not consistently apply the DFD grammar, and instead incorporated “object-like” components into the DFD. It was also observed that the same “object-like” structures showed up in one of the Text participant's diagrams who specifically set out to create a DFD. These result imply that participants gravitated towards “object” grammars regardless of which treatment they were given, and that the variance in diagrams within the OOA participants was small relative to the other treatment groups.

It should be recognized, however, that these claims that objects are more “natural” cognitive structure may be a result of the description provided to the participants. In other words the way the text is formulated and described may naturally lead individuals to an object approach. Objects may be natural for the case considered, but not natural for other cases described in a different way. [13]

So what do these observations enable us to say? Given the possible bias of the case descriptions the conclusions are listed below:

The observations in the protocol analysis indicate that for even simple problems the connection between DFD and ERD are difficult for individuals to make. Further, novices tend to gravitate towards "object-like" constructs when reasoning about a system. And when provided with an OOA model, participants tend to organize and refer to the OOA more often that a DFD/ERD combination. Together these observations suggest that the object model may be a more "natural" approach to system analysis than the separation of data flows from data structure recommended by structured analysis. Object oriented analysis has the potential, therefore, to improve how individuals reason about and come to an understanding of a domain.

7.6 Hypothesis H2.1: Optional vs Mandatory

Chapter 4.2.1 outlined an argument using the BWW (Bunge–Wand–Weber) ontology to suggest that an entity relationship model grammar that required mandatory properties with subtyping would be more understandable than a grammar that allowed the use of optional properties. The justification for the argument is that both the optional and mandatory grammars can be shown to possess the same level of ontological completeness but the mandatory grammar with subtyping has more ontological clarity than the optional grammar. The greater ontological clarity associated with the mandatory grammar will make the mandatory grammar less ambiguous and easier to understand than the optional grammar. It is argued, therefore, that individuals viewing ER diagrams created using the mandatory grammar will display a higher level of understanding than individuals viewing ER diagrams created with the optional grammar.

This argument has been outlined in Bodart & Weber (1996) and Bodart et. al. (1998). These papers also outline three empirical attempts to evaluate the differences in understanding between thee two groups. The first two of these tests are discussed below as a preliminary to the results developed in this study. The third experiment, which was a replication of the tests the researcher developed for this thesis, will be reported at the end of this section.

The first experiment describe in Bodart & Weber (1996) utilized a "free-recall" task that showed a diagram to participants for a three minute interval and then asked participants to redraw the diagram from memory. In this experiment the optional group outperformed the mandatory group across a variety of measures. Bodart et. al. (1998) note that this result was

[13] I would like to thank Carson Woo for introducing this point.

opposite to the result predicted and contradicted the mandatory versus optional argument. This result is not surprising, however, when viewed in light of Mayer's (1989) experiments.

Mayer suggested that individuals interpret and modify information they receive in order to create a level of understanding. The more an individual understands from a diagram or text description, the less likely he or she is to store that information in verbatim form. In other words, if a person simply memorizes, but does not understand, then the information is stored in verbatim form. When an individual creates understanding, the information is more likely to be modified. Since the free-recall tested primarily the "memory" of individual, the first experiment reported in Bodart et. al. (1998) was not a true assessment of understanding. It is not surprising that the optional group scored higher in recreating the diagram for two reasons. First, the optional diagram had less "things" to remember. This is a direct result of the increased ambiguity inherent in the optional model. There are less "things" to remember in the optional diagram, but the "things" can be interpreted in a number of ways. Second, if the mandatory diagram is more easily understood, then the mandatory diagram would be ***less*** likely, according to Mayer, to be remembered verbatim. This would confirm the results in Bodart et. al. (1998). Since it would be impossible to separate the influence of these two effects noted above, we cannot conclude from this experiment, so it is impossible to state that participants viewing the mandatory model had more understanding that the optional participants. What can be said, however, is that the first result reported by Bodart et. al. (1998) do ***not*** contradict the prediction that participants viewing mandatory diagram develop a higher level of understanding than participants viewing optional diagrams.

The second experiment reported in Bodart et. al. (1998) substituted the "free-recall" measure for a set of comprehension questions. The ten comprehension questions were focused on details of the diagram and required the participants to answer "yes", "no", or "not sure". Two cases were assessed. The results from the experiment showed there were no significant differences in accuracy or time between the optional and mandatory groups. In the second case, the optional group was found to perform better in the accuracy than the mandatory group, but not significantly different in terms of time. The average difference between the mandatory and optional group in terms of accuracy was approximately 1 point out of 10 in the second case, whereas in the first case it was 0.80.

Bodart et. al. (1998) report that these *finding "did not support and indeed were contrary to the hypothesis that motivated our work."* Again, in the light of Mayer's experiments, it could be argued that these results could have been anticipated. Mayer (1989) notes that if two representations provide the same basic information to a participant, then the treatment groups are not likely to score differently on comprehension questions, unless the questions are focused on conceptual understanding. Since the questions used by Bodart et. al. (1998) were focused on aspects within the diagram and not on the conceptual understanding of the diagram, then Mayer would predict that there would be very little difference between the accuracy of the two treatment groups. In general, this was the result that Bodart et. al. (1998) found. It is important to note that these results do ***not*** contradict the hypothesis that the mandatory group will develop a higher level of understanding than the optional group. Instead, these results can be viewed as an affirmation of Mayer's prediction that the treatment groups will not score significantly different when asked comprehension questions.

The results provided by Bodart et. al. (1998) set the stage for the intragrammar comparison outlined earlier in thesis in Chapter 4 and 5. The empirical procedures in the current study differ from Bodart & Weber (1998) and the first two experiments in Bodart et. al (1998). The main differences are the use of "problem solving" questions as suggested by Mayer (1989), and the inclusion of a Cloze test. The argument developed by Bodart & Weber (1996) and noted above has led to the following hypothesis:

Hypothesis P2.1

P2.1: ERD's that use a grammar requiring subtypes with mandatory attributes will promote a higher level of understanding in persons viewing the diagram than ERD's developed using a grammar allowing the use of optional properties.

On the basis of this argument, predictions regarding the relative performance of the "mandatory" and "optional" grammars can be made. We will refer to five predictions:

1. **Comprehension scores for individuals provided with ERD's created using an "optional" grammar will NOT be significantly higher than individuals provided with ERD's created with a "mandatory " grammar**. Given the relatively small differences between the constructs in the two grammars, the differences in conceptual information presented in the diagrams is likely to be small. This suggests that the comprehension scores across the optional and mandatory groups will not differ significantly. This result is likely for most intragrammar comparisons.

2. **Problem solving scores for individuals provided with ERD's created with a "mandatory" grammar will be significantly higher than individuals provided with ERD's created with a "optional" grammar**. This prediction follows from the argument that the mandatory grammar produces a diagram with a higher ontological clarity than the optional grammars. This ontological clarity reduces the ambiguity associated with the ERD. The optional grammar, being less ontologically clear,

introduces ambiguity into the ERD and will result in less well-formed cognitive models of the system domain. The higher ontological clarity associated with the mandatory grammar can promote, therefore, a higher level of understanding among participants, which will result in higher scores on the problem solving test.

3. **Cloze scores for the participants provided with ERD created using the mandatory grammar will be higher than participants provided with ERD's created using the optional grammar.** Since neither "mandatory" participants nor "optional" participants will have seen the original text description, the Cloze test can act as a test of the overall understanding of the domain. The argument that the mandatory grammar will promote a "better" cognitive model of the domain than the optional grammar suggests that Cloze test scores should be higher for participants provided with mandatory descriptions.

4. **No difference in the perceived ease of use scores will be observed between participants provided with ERD's created from either the mandatory or optional grammar**. This prediction follows from the argument made in prediction 1 above. The similarity between the ERD's created by the two methods suggests that individuals would find little difference in the perceived ease of use between the two methods.

5. **The time taken to complete the comprehension, problem solving, and Cloze test will be shorter for participants using ERD's created using the mandatory grammar**. This prediction again relies on the argument that the ontological clarity associated with the mandatory grammar will reduce the ambiguity introduced during the interpretation process. The reduced ambiguity associated with the mandatory grammar will result in less time sorting out the ambiguities and hence less time required to answer comprehension questions and fill in the Cloze test.

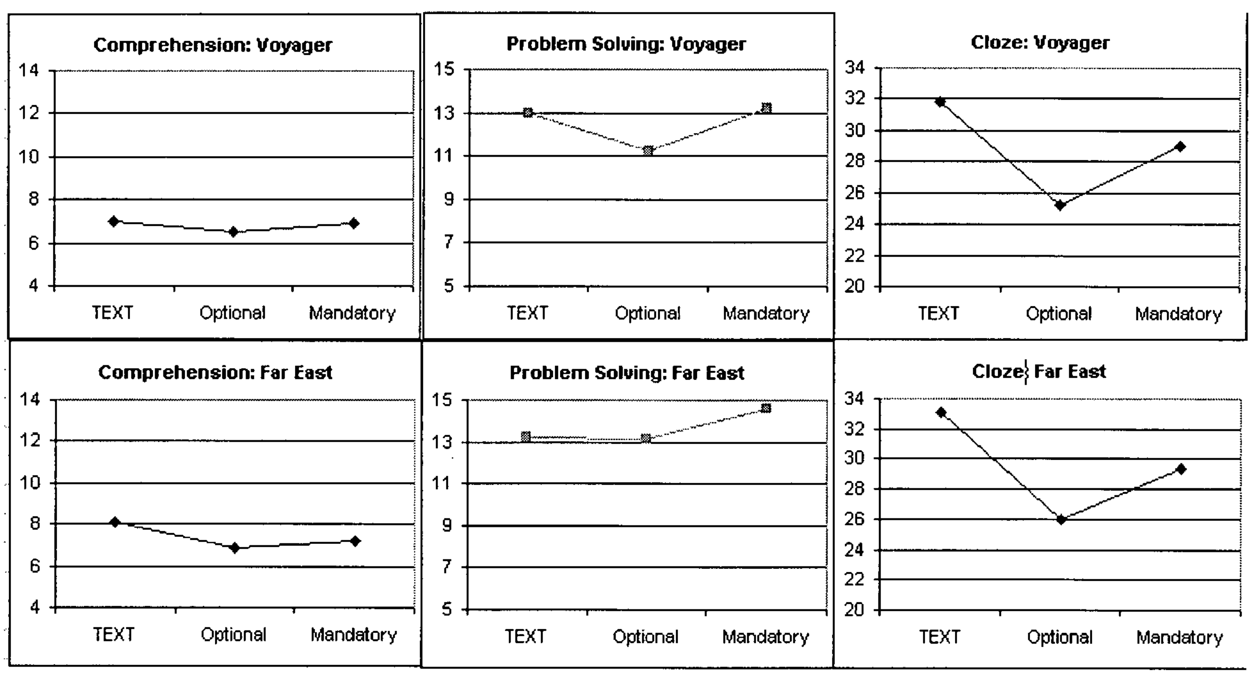
To assess these predictions we used MANCOVA analyses using the data from the intergrammar study. A short review of the analysis from the previous analysis will refresh our understanding of the MANCOVA applied here. Two separate MANCOVA analyses were performed, one related to the VOYAGER case and one related to the FAR EAST case. Within each case we analyzed three product measures and the four variables related to effort. Before seeing the results of the MANOVA, the means and standard deviations for each of the treatment groups is provided in Table 41 below.

Table 41: Means Across Treatment Groups for Dependent Variables

Dependent Measure	Case: VOYAGER			Case: FAR EAST		
	TEXT	Optional	Mandatory	TEXT	Optional	Mandatory
COMP	7.00 (1.77)	6.55 (1.38)	6.89 (1.42)	8.12 (2.49)	6.84 (1.91)	7.17 (2.32)
PROB	12.98 (4.42)	11.28 (4.15)	13.20 (3.41)	13.21 (4.96)	13.10 (2.67)	14.58 (3.19)
CLOZ	31.68 (5.05)	25.21 (6.17)	29.00 (6.08)	33.15 (5.97)	26.00 (5.80)	29.38 (5.08)
COMP TIME	10.15 (8.86)	9.93 (9.49)	10.55 (8.84)	12.73 (8.17)	12.44 (8.55)	12.71 (9.10)
PROB TIME	19.31 (7.54)	16.78 (7.48)	16.42 (8.48)	24.75 (8.87)	20.53 (9.88)	20.25 (8.21)
CLOZ TIME	8.88 (4.19)	10.07 (3.59)	9.39 (2.86)	8.97 (4.28)	10.82 (4.75)	10.17 (3.89)
Ease of Use	2.042 (2.48)	-0.414 (1.96)	-1.27 (1.91)	2.152 (2.52)	-.383 (2.06)	-1.46 (1.95)

Having viewed the means, we can graph the means scores for the problem solving and across the treatment groups. These are shown below in Figure 24.

Figure 24: Means for Problem Solving and Comprehension :Far East and Voyager



The graphs shows a consistent theme, with little or no difference between groups in comprehension, but some difference emerging between the “Mandatory” and “Optional” groups on the problem solving questions. The fact that the differences are smaller than in the previous intergrammar comparison, is somewhat predictable, as there is less variance in the grammars associated with intragrammar comparisons. Having looked at the graphs, we can now turn our attention to the MANOVA results. The MANOVA results are displayed in Table 42 below.

Table 42: MANCOVA for Hypothesis P2.1

Item	Case: Voyager		Case: Far East	
1. Bartlett's Sphericity				
	Chi	**Sig.**	**Chi**	**Sig.**
Are dependent Var.correlated?	137.5	.000	77.56	.000
2. Box's M				
	M	**Sig.**	**M**	**Sig.**
Homogenous covar. Matrices?	72.85	.000	81.75	.000
3. Within Cell Regression Effect				
Covariates significant?	λ	**Sig.**	λ	**Sig.**
KnowMeth	.005	.910	.015	.682
KnowIFIP	.017	.639	N/A	N/A
KnowEVEN	N/A	N/A	.016	.640
4. Factor*covariate Interaction Effect				
	λ	**Sig.**	λ	**Sig.**
Homogenous regress. effects?	N/A	N/A	N/A	N/A
5. Main Effect				
	λ	**Sig.**	λ	**Sig.**
Significant effects between methods?	.554	.000	.540	.000

The results in Table 42 are similar to the result provided for Hypothesis H1.1. The Bartlett's tests of sphericity indicate that the MANOVA procedure is appropriate for both cases as the dependent variables show a significant amount of multivariate correlation. The Box's M test indicates that the assumption of homogeneous covariance matrices is violated in the MANCOVA. As discussed in Appendix J, the effect of the violation is to decrease the power of the test. Type I error rates are affected only slightly except in extreme cases, and there is nothing to indicate that the conditions in this case are extreme.

The covariates in both MANCOVA procedures were not significant. This finding was to be expected given the very low correlations observed between dependent measures and the covariates. Since the covariates were not significant, the main effects could be estimated without including the effect of the covariates. The covariates were dropped, therefore, from all four of the models. Dropping the covariates eliminated the need for the factor*covariate interaction effect, which was not reported.

The main effects show strong significance in both of the MANCOVA procedures. These findings suggest that there are significant differences among the treatment groups. The finding also suggests that these differences were observed across both cases. While these finding are encouraging, it is necessary to perform a univariate post hoc analysis on the dependent variables to see which variables contributed to the observed differences.

The univariate tests are one-way ANOVA procedures that produce an F statistic. One ANOVA procedure is produced for each dependent variable. For each case in the study, seven dependent measures - three product and four process measures - need to be considered. When an ANOVA test shows significant results, it indicates that the variable being analyzed contributed towards the differences observed in the multivariate procedure. The univariate tests provide a method for addressing the predictions made in Table 31 at the start of this chapter. These univariate tests are, therefore, an important part of the analysis. The univariate findings are provided in Table 43 below. Significant results are highlighted in gray.

Table 43: Univariate Results for Hypothesis P2.1

Item	Case: VOYAGER			Case: FAR EAST		
6. Univariate Results						
Which variables contribute to differences?	**F**	**Sig.**	**Observed Power**	**F**	**Sig.**	**Observed Power**
Comprehension	0.85	.430	.193	2.969	.056	.566
Prob. Solving	2.53	.046	.598	2.263	.109	.451
Cloze	11.24	.000	.991	13.90	.000	.998
Comp. Time	0.046	.955	.057	0.089	.915	.063
Prob. Solv. Time	1.356	.262	.257	3.045	.052	.578
Cloze Time	.990	.375	.218	1.479	.233	.310
Ease of Use	21.49	.000	1.00	21.38	.000	1.00

The results in Table 43 above indicate that we observe significant differences between treatment groups in the Cloze test, ease of use, problem solving (for one case), and comprehension and problem solving time. These ANOVA tests indicate significant differences among the three treatment groups. The ANOVA procedures do not indicate in which direction the differences lie, nor do they report on the pair-wise comparisons between two techniques. In developing and analyzing pair-wise comparisons and related significance tests, it is important to begin with the means for each of the three treatment groups. Table 44 below provides these means for the dependent measures across the three treatment groups.

To determine the pair-wise comparisons, a set of simple contrasts was created. Recall, that simple contrasts involve comparing the estimated means - which may differ slightly from the actual means shown in Table 44 - to a chosen "control" group. For Hypothesis P2.1 we chose the "Mandatory" group as the "control" group for the contrast. Two "Simple" contrasts are created; One contrast comparing Mandatory and Optional and the other comparing mandatory and text.

The results of these contrasts for each of the dependent variables over the two cases, are provided below in Table 44 and 45. Table 44 includes results for the product variables and Table 46 displays results for the process variables. Significant contrasts are highlighted in gray and significance levels are calculated using the Bonferroni correction. Note that the difference ("Opt. – Mand." or Text – Mand.") represents the estimated difference between means. A negative difference, therefore, suggests that the Mandatory group outperformed the other group (optional or text) on that variable for that treatment group.

Table 44: Contrasts for Optional vs. Mandatory Comparison (Product Variables)

Contrast	Item	Case: VOYAGER			Case: FAR EAST		
		COMP	PROB	CLOZ	COMP	PROB	CLOZ
Optional and Mandatory	Opt. – Mand.	-.345	-1.96	-3.789	0.937	-1.687	-3.43
	Significance	.324	.038	.005	.086	.051	.010
Text and Mandatory	Text - Mand.	0.130	-0.268	2.78	-0.365	-1.433	3.77
	Significance	.779	.779	.048	.491	.105	.006

Four effects were consistent across the 2 cases. As predicted, the Mandatory group scored significantly higher than the Optional group on the problem solving. It is interesting to note that the differences between the Mandatory and Optional groups are smaller than those observed in the previous study between the OOA and DFD/ERD. The smaller differences were expected between the Mandatory and Optional groups as there are fewer differences between the grammars in this intragrammar study than in the previous intergrammar study.

The Mandatory group also outperformed the Optional group on the Cloze test in both cases. This consistent result was not observed in the intergrammar study. A closer, qualitative look at the individual items in the Cloze test revealed that the difference in Cloze test scores was not the result of a few words consistently missed by the Optional group and correctly answered by the Mandatory group. Instead, the higher Cloze score was the result of a slightly higher average in many of the items. This result indicates that for the intragrammar assessment, the Cloze test is capable of detecting differences even when the comprehension scores are not significantly different. This suggests that the Cloze test is slightly more sensitive to differences in understanding than the 'yes-no-unsure" comprehension test, and somewhat less sensitive than the problem solving test.

Another consistent result from Table 44 above was the Text group consistently outperformed both the Mandatory and Optional group in the Cloze score. This is natural as the Text group was exposed to the text description before taking the Cloze test. This result again adds to the internal validity of the study by indicating that the Cloze test was not a simple exercise, and that individuals scored differently based on the information that was presented to them. In other words, the Cloze test was not simply an assessment of general knowledge, but an assessment that required some specific domain knowledge.

The final result from Table 44, as seen earlier in the intergrammar study, is that no consistent difference was observed across the two cases in regards to the comprehension test. The lack of significant differences on the 12 question comprehension task, across two studies, again suggests that the three presentation methods were roughly equal in their ability to transmit the "content" of the domain to the participant viewing the diagram or description. The claim that the information embedded in one or more of the presentation methods gave an advantage to one or more of the groups is not substantiated by the findings. This finding also provides further justification for Mayer's (1989) claim that small differences in comprehension can be observed at the same time as significant differences in problem solving are observed.

The four observations discussed above provide an indication that the Mandatory group was able to develop a “higher” level of understanding than the Optional group in the two cases. The findings also help to establish that the instrument, proposed in this thesis, is sensitive enough to pick up differences in an intragrammar study. We have been able to differentiate between the presentation methods in all four cases using the problem solving questions. These findings indicate, therefore, that the instrument proposed in this thesis is relatively robust.

Having addressed the major findings with regards to the product measures, we can now move to address the measures of "effort". Table 45 presents the results of the comparisons between the Mandatory “control” groups and the other Optional and Text groups. Significant contrasts are highlighted in gray and significance levels are calculated using the Bonferonni correction.

Table 45: Contrasts for Optional vs. Mandatory with Effort variables

Contrast	Item	Case: VOYAGER				Case: FAR EAST			
		COMP Time	PROB Time	CLOZ Time	EASE	COMP Time	PROB Time	CLOZ Time	EASE
Optional and Mandatory	Opt. – Mand.	-.620	2.88	-.519	0.85	-.1.06	-.038	.594	.873
	Significance	.765	.128	.542	.086	.692	.985	.562	.090
Text and Mandatory	Text - Mand.	-.396	.359	.671	3.31	-.175	4.51	-1.20	3.307
	Significance	.895	.842	.409	.000	.949	.032	.250	.000

Two observations regarding Table 45 are interesting. First, the Text group found the text description significantly easier to use than either the Mandatory or Optional diagrams. This result was also observed in the intergrammar comparison. This observation lends credence to the argument that participants may perceive text descriptions as easier to use because the participants use text so often. This finding should not be viewed as a call for eliminating the use of optional properties in ERD. Instead, the results suggest that, wherever possible, to separate schema definitions when assigned properties change the way that an entity (or relationship) is viewed. To take the example in chapter 4, if a pager number indicates a new tretmant for the "Customer" entity, split the entity into two sub-types and assign mandatory attributes accordingly.

The second observation from Table 45 is the fact that there were no significant differences in the time spent on the three study tasks (comprehension, problem solving, and Cloze). This "non-finding" may be the most important statistic in the table. The insignificant differences observed in the table suggest that the Mandatory and Optional treatment groups spent similar amounts of time on the comprehension, problem solving, and Cloze tasks. If this is true, it is then possible to state that the improved scores associated with the Mandatory group on the problem solving and Cloze tasks are not directly related to the time applied to the task. In other words, the higher scores in problem solving and Cloze test are not the result of taking more time. If we assume, as Mayer (1989) argues, that the level of understanding is directly related to problem solving scores, then we can suggest that the diagram created from the Mandatory grammar is able to communicate a deeper level of understanding of a domain, in the same amount of time, as the diagram created with the Optional grammar.

Conclusions from Intragrammar Study

Combining the observations from Table 44 and 45 enables us to make the following statement:

The data collected in this intragrammar study provide some evidence for the conclusion that ER diagrams created with the "Mandatory" grammar communicate at least the same amount of comprehensive knowledge, and a measurably higher level of understanding regarding a domain, in the same amount of time, as diagrams created with the "Optional" grammar.

The study, therefore, has provided evidence that supports Hypothesis P2.1 and most of the predictions made as a result of this hypothesis. The findings provide some evidence for the assertion that grammars that have a higher level of ontological clarity, measured in terms of the BWW ontological constructs, provide a greater opportunity for understanding than grammars with less ontological clarity. As discussed early in chapter 4, the results suggest the use of mandatory properties over optional properties when feasible. The ERD created with mandatory properties seem to reduce the ambiguity associated with collapsing separate schema definitions as happens in the use optional properties.

The results of the intragrammar study should not be viewed as an overall condemnation of the use of optional properties. In interpreting these results, it is important to separate conceptual modeling from database design. The results suggest that in conceptual modeling, the use of subtypes rather than optional properties can significantly improve understanding. Taken to the extreme, however, mandatory grammars might lead to a complex model and an unmanageable database design due to proliferation of types. Thus, the use of mandatory or optional properties remains a question of balance between the reduced ambiguity of subtypes with mandatory attributes and simplicity provided by the optional grammar. The results in this study suggest that the mandatory grammar should be considered carefully for conceptual modeling.

It should be noted that in developing the hypotheses for the intragrammar comparisons, the comparison with text was not analyzed in detail. It is interesting to note that the text and mandatory groups displayed similar abilities in all three of the tests. This finding is significant since text was found to be significantly easier to use for novice users. The findings suggest, therefore, that text may be as good as diagrams in communicating knowledge of conceptual data structures. This result requires more analysis and will be focused on in future research.

Turning back to the Bodart et. al. (1998) study, the third experiment they performed is of interest in the light of our finding. The description of the experiment from Bodart et. al. (1998, p. 20) is provided below:

> *"The third experiment was motivated by a personal communication we received from Gemino notifying us of results he had obtained from research he had undertaken as part of his Ph.D... Like us, Gemino proposed that conceptual models that used subtypes with mandatory properties would communicate more meaning to their users than those that used optional properties. Like us, also, Gemino used a "verbatim" recall test and a comprehension test to evaluate the performance of participants in the experiment he undertook. Unlike us, however, Gemino also had his participants undertake a problem-solving task. Furthermore, he measured his participants' perceptions about the ease of use of the two representations."*

The experiment was tested using two cases: the "Voyager case which was used in the intragrammar study, and a "University" case. The experiment essentially replicated the study described in this chapter except for two differences. Participants were limited to 10 minutes in the comprehension test and 30 minutes in the problem solving test in the Bodart study rather than having unlimited time to accomplish the tasks. The time limits in the comprehension test may have been a constraint as the average time to complete was over 10 minutes for almost all of the groups. The 30 minute problem solving time limit was well above the mean times reported earlier in this chapter.

The results of the test are provided below. Since the Voyager (bus-route) case was replicated, the tests for this case should be comparable. The report of the results is again taken from Bodart et. al. (1998, p. 26):

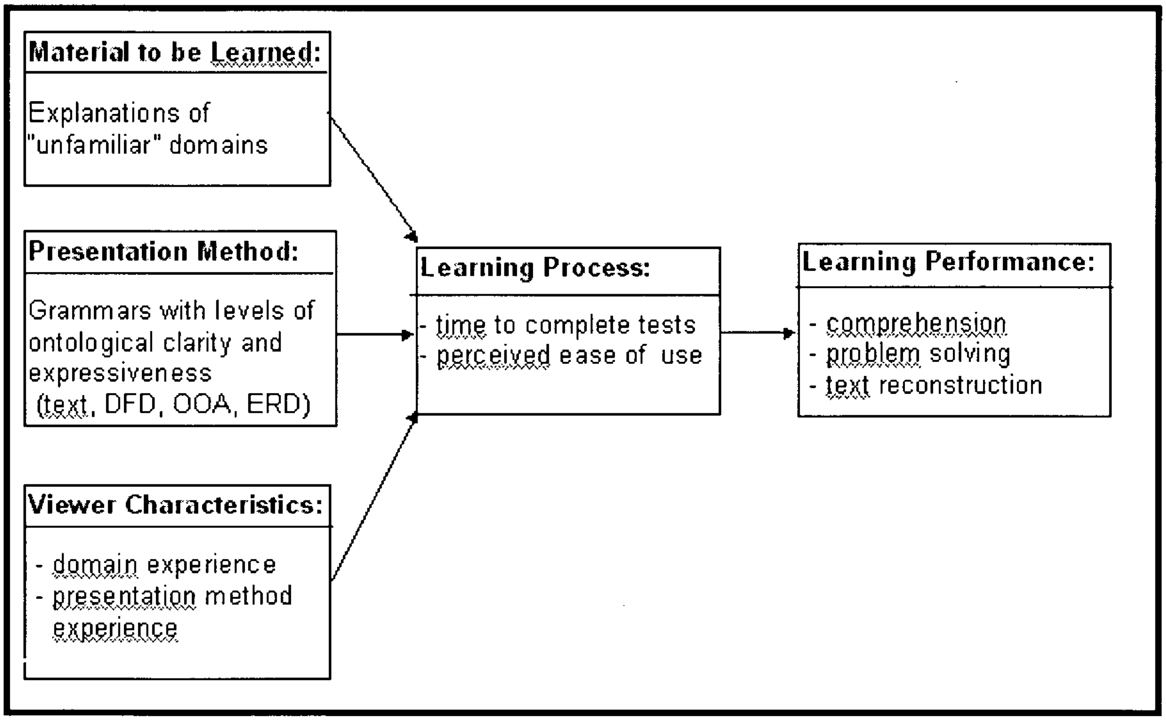
> *"Problem-solving performance: For both the bus-route domain and the university-research domain, the mandatory-properties group outperformed the optional-properties group in terms of the number of correct answers based on the entity-relationship diagram...*
> *Comprehension Performance: For both domains, there were no differences between the performance of the two groups in terms of comprehension performance.*

These results extend the findings to three different cases and two separate research teams. The Bodart et. al. (1998) study, therefore, provides a valuable external validation of the results obtained and reported earlier in this chapter. This section concludes the major findings in this study. In the following section we will look briefly at the impact that previous knowledge of analysis methods and prior domain knowledge had on the both the intergrammar and intragrammar studies.

7.7 Hypothesis 3.1: The Effects of Learner Characteristics on the Findings

We argued, in Chapter 3, that the process of interpreting a diagram created with a system analysis technique was a learning process. Three antecedents to the learning process were identified: the material to be learned; the method used to present information (presentation method); and the characteristics of the learner (observer). An overview of the learning process is provided below in Figure 25. The experiments described in this thesis treated these antecedents in three different ways. The material to be learned, for example, was held constant across all participants in a study. The diagrams were created from the same text description so that the content was - in general – similar across all participants. The presentation method was the treatment variable in each of the studies and was strictly controlled, therefore, by the researcher. In contrast, the characteristics of the learner were not tightly controlled. Instead, the randomization of participants into treatment groups was used to control for differences in domain or analysis method knowledge between treatment groups. A pretest was used to collect the knowledge of system analysis methods and domain knowledge for each participant. These two measures were then used as covariates in the MANCOVA analysis. When used in this way, the effect of previous knowledge of either analysis methods or domains would be factored out and the power of the test would be increased.

Figure 25: Overview of Elements in the Learning Process



The results reported in this chapter indicate that the presentation method is an important factor in the determining the understanding that is developed by individuals interpreting a description of a domain. These results were the main treatment effects reported earlier for both the intergrammar and intragrammar study.

The results from the experiments also indicate that the level of domain knowledge and level of knowledge of analysis methods were NOT significant factors in estimating the level of "understanding" developed by participants in the study. The effect of the covariates were estimated using two covariates in each MANCOVA analysis and are reported in Tables 32 and 45 above. These results were predicted in Table 21 of Chapter 6 when the correlation matrices between dependent variable and the covariates indicated there was little or no linear relationship between the covariates and the dependent variables.

The argument for the effect of prior method knowledge on the level of understanding is straightforward, perhaps even obvious. The more experience and competence a person has with regards to a particular analysis method, all other things equal, the better the learning outcome when the analysis method is used to describe a domain. The same is true for domain knowledge. The more domain knowledge a person has prior to viewing a diagram or description, the more likely that the person will develop a higher level of understanding, all other things equal. This argument led to the following Hypothesis:

Hypothesis P3.1

P3.1 Individuals with more experience and competence with a presentation method used to represent a domain, will develop higher levels of understanding when compared with individuals who have less experience and competence with a presentation method. Similarly, individuals with more domain specific knowledge will develop higher levels of understanding when compared with individuals who have less domain knowledge.

While the argument seems obvious, the results in both the intergrammar and intragrammar comparisons did not confirm this hypothesis. In fact, there was not a single MANCOVA that showed even a marginally significant effect ($\alpha = 0.10$) from either the level of domain knowledge or analysis method knowledge.

Several reasons can be offered to explain this surprising conclusion. It should be recognized that the cases used in the studies may have been simple enough to understand without significant prior domain knowledge. This would explain the insignificant effects of prior domain knowledge. The cases may also have been familiar enough to the participants so that may not have been needed knowledge of the analysis methods to understand the diagrams. This would explain the insignificant effects of analysis method knowledge for those participants using diagrams, but not for those participants using text. Another reason may be that the relationship between prior knowledge of the domain and analysis methods may be non-linear. MANCOVA and correlation are measures of linear relationship and may not be good estimates of the actual relationship between the variables.

Perhaps the most plausible reason for the insignificant results is measurement error. The presence of measurement error is likely as empirical researchers in systems analysis have not developed a set of standardized instruments for prior domain knowledge or knowledge of analysis methods. The lack of standardized instruments would explain the insignificant results as the measures proposed in this thesis have not been verified in other studies. This is particularly true of the measure of domain knowledge as noted earlier in section 6.2. The low reliability associated with the domain knowledge measure coupled with the lack of a verified instrument indicates that any conclusions regarding the effect of domain and analysis method knowledge in this study should be viewed with a significant level of skepticism.

This analysis and the results observed previously in both studies enable us to make the following statement regarding the Hypothesis H3.1 above

Conclusion

The data collected in the intergrammar and intragrammar studies provides evidence for the conclusion that neither the observed level of prior domain knowledge nor the observed level of analysis method knowledge had a significant effect on the level of "understanding" that participant's developed regarding a domain. This conclusion should be viewed with skepticism, however, due to the lack of valid and verified measures for domain and analysis method knowledge.

It appears obvious to the researcher that prior domain knowledge and experience with analysis methods, all other things equal, will help individuals to develop a deeper level of understanding of a domain. It is important to note, therefore, that the conclusion stated above holds only for the test performed and should not be viewed as a general statement regarding impact of domain and analysis method knowledge.

Further Considerations of Learner Characteristics

To this point in the analysis, learner characteristics have been defined in terms of prior knowledge of either the domain or the methods used to describe the domain. There is more to participants, however, than just these two characteristics. For example, an important consideration in the characteristics of the observer is the "cognitive style" that an individual brings to the task[14].

Another important consideration is the curious results observed in Chapter 6 regarding the correlation between common task in different cases. For example, in the intergrammar study, a high correlation (0.626) was found between the problem solving scores in the EVENT case and the problem solving scores in the IFIP case. The same was true for the intragrammar study. The correlation results originally reported in Tables 19 and 26 from Chapter 6 are recreated below in Table 46 and 47.

[14] Special thanks to Genevieve Basselier, Yair Wand, and Izak Benbasat who independently suggested this idea.

Table 46: Dependent Variable Correlation Coefficients (Intergrammar Study)

	Case: Event			Case: IFIP		
Variable	EVCMPTOT	EVPRBTOT	EVCLZTOT	IFCMPTOT	IFPRBTOB	IFCLZTOT
EVCMPTOT	1.00 p=.000	X	X	X	X	X
EVPRBTOT	-.099 p=.311	1.00 p=.000	X	X	X	X
EVCLZTOT	.084 p=.392	.2170 p=.025	1.00 p=.000	X	X	X
IFCMPTOT	**.268 p=.006**	.001 p=.989	.191 p=.049	1.00 p=.000	X	X
IFPRBTOB	.136 p=.171	**.6264 p=.000**	-.022 p=..817	.095 p=.329	1.00 p=.000	X
IFCLZTOT	.047 p=.628	.147 p=.132	**.649 p=.000**	.129 p=.188	-.056 p=.565	1.00 p=.000

Table 47: Dependent Variable Correlation Coefficients (Intragrammar Study)

	Case: Voyager			Case: Far East		
Variable	EVCMPTOT	EVPRBTOT	EVCLZTOT	IFCMPTOT	IFPRBTOB	IFCLZTOT
EVCMPTOT	1.00 p=.000	X	X	X	X	X
EVPRBTOT	.005 p=.955	1.00 p=.000	X	X	X	X
EVCLZTOT	.001 p=.995	..039 p=.686	1.00 p=.000	X	X	X
IFCMPTOT	**.391 p=.000**	.178 p=.064	.118 p=..221	1.00 p=.000	X	X
IFPRBTOB	-.069 p=.479	**.504 p=.000**	-.099 p=.308	.020 p=.836	1.00 p=.000	X
IFCLZTOT	.117 p=.226	.16 p=.083	**.708 p=.000**	.188 p=.050	-.067 p=.491	1.00 p=.000

The strength and consistency of the results in Table 46 and 47 above suggest that there is more to these observations than simple a random occurrence. Most important, ***the correlations are observed primarily between the same task across different cases and***

not between different tasks within the same case. The positive correlations indicate that participants, for example, who scored high the problem solving task in the first case, were likely to score high in the problem solving task in the second case. And further, the correlation ***between*** cases was higher than ***within*** a single case. This means that a participant scoring high on problem solving in the first case was less likely to score high in the comprehension or Cloze test ***within the same case*** than he or she was to score high on the problem solving task in the second case.

These results indicate that when a participant views a diagram or text description, the participant also bring along his or her relative abilities to complete the comprehension, problem solving, and Cloze tasks. A participant who scores high in problem solving in the first case, for example, may score highly because he or she has problem solving task "skills". Regardless of the case, this type of individual will score higher of problem solving than other individuals because of these skills. The same holds true for participants with Cloze "skill" or comprehension "skill."

While randomization of participants would be expected to eliminate significant differences across groups in the problem solving, comprehension, and text reconstruction "skills", the result is still of interest. If problem solving "skills" do exist, and problem solving is a measure of a "higher" level of understanding, then it appears the possibility exists for training individuals to improve their use of problem solving skills. This suggests that researchers should not only consider how an analysis method presents information, but also how the observer approaches and thinks about the information. Perhaps by asking observers the right types of questions, and encouraging them to think about the information in a different way, system analysts may improve the "understanding" that individuals develop regarding a system. These questions are left to a future study.

7.8 Summary of Results

This chapter has presented results of the empirical research conducted in this thesis. The first two sections of the chapter outlined the framework for analysis of the MANCOVA results. This framework was then used as a foundation of the Hypothesis test found within the chapter.

The third section provided results from the MANCOVA that compared the "understanding" developed by participants using either graphical models (OOA, DFD/ERD) or text descriptions. The results provided evidence that participants provided with graphical models scored higher in problem solving with no significant difference in comprehension scores. A difference was also observed between the type of graphical models provided to the participants. This indicated that not all graphical models provide the same advantages over text, which was an important initial finding for the subsequent Hypothesis.

Section 7.4 reported the results from the intergrammar comparison of OOA and the combination of DFD/ERD. The results showed that participants using the OOA diagram scored significantly higher in problem solving with no significant differences in comprehension or text reconstruction (Cloze) scores between the two groups. The following section, Section 7.5, reported on a search for reasons why these results were observed. This section explored protocol data and diagrams created by participants. The protocol analysis provided three results. First, participants provided with the DFD/ERD combination did not make substantial use of the ERD. Second, the participants provided with OOA were more likely to refer to the OOA diagram constructs and to use the diagram constructs as a conceptual "anchor" to search for answers to the problem. Finally, when drawing their own representations, participants consistently chose to use "object-like" constructs, regardless of whether they originally saw the OOA or not. These "object-like" constructs were apparent even when participants originally set out to draw a DFD. These results suggested that objects may be a more natural cognitive structure than those provided in the DFD/ERD combination.

The attention in Section 7.7 was then turned to the intragrammar comparison of mandatory and optional properties in the Entity Relationship Diagram. The results of the MANCOVA analysis indicated that participants who were provided with diagrams created using the Mandatory grammar scored higher in problem solving and text reconstruction tasks than the participants provided with diagrams created from the Optional grammar. In addition, the differences observed between the two ERD grammars were smaller than those observed between the OOA and DFD/ERD comparison. The smaller differences were anticipated as the two grammars in the intragrammar study were more similar than the OOA and DFD/ERD in the intergrammar study. This result provided a further indication that the instrument developed in this thesis to compare analysis methods was performing as predicted.

Section 7.6 combined the results in both the intergrammar and intragrammar studies to analyze the effect of learner (observer) characteristics on the level of "understanding" developed by participants. The results showed that prior knowledge of either the domain or the analysis methods used to describe the domain did not have a significant effect on the level of understanding developed by participants in either the intergrammar or intragrammar study. These results were viewed with some skepticism largely due to the absence of a verified standard instrument for measuring prior domain knowledge or knowledge of analysis methods.

In the next chapter we will summarize the conclusions reached in Chapter 7 and revisit the objective of the thesis stated in Chapter 1. After discussing these conclusions, some consideration of the internal and external validity of the results will be noted, some limitations of the study will be discussed, and suggestions for future research are addressed.

Chapter 8: Conclusions

The objectives of this thesis were outlined in Chapter 1, section 1.3. Two objectives were identified. The first was to develop an empirical instrument and procedure for comparing systems analysis modeling techniques, and more specifically, the grammars that underlie these analysis techniques. The second objective was to apply this comparative empirical instrument and procedure to evaluate actual grammars. Two of these studies were undertaken. The intergrammar study drew a comparison between text description (TXT), Object Oriented Analysis (OOA), and a combination of Data flow Diagramming (DFD) and Entity Relationship Diagramming (ERD) techniques. The intragrammar study compared different grammars suggested for the Entity Relationship Diagram (ERD).

The success of the thesis can be measured, therefore, in two ways. Of primary interest is the strength of the empirical instrument and procedure developed in Chapter 3, 4 and 5. If validity and reliability of the instrument and procedure can be established, then the results of the two studies are of interest. The discussion of conclusions, therefore, begins with a discussion of the instrument.

8.1 Conclusions Drawn Regarding the Instrument

To establish the "strength" of an instrument, a discussion of the validity of the instrument is essential. Straub (1989) noted the importance of instrument validation in MIS research. Straub noted that validation 1) introduces an element of "rigor" into the research methodology; 2) promotes cooperative research efforts by enabling confirmatory research[15]; 3) leads to an "improved clarity to the formulation of research questions"; and 4) reduces the uncertainty associated with unconfirmed findings. Establishing the validity of an instrument is a difficult task. A framework for the discussion of validity related to an instrument is provided in Cook and Campbell (1979). The framework identifies four areas that threaten the validity of an instrument: 1) construct validity; 2) internal validity; 3) statistical conclusion validity; and 4) external validity. These four areas are addressed below.

Before discussing the four areas of validity suggested by Cook & Campbell (1979) it is necessary to refer to the point made by Straub (1989, p. 150):

---

[15] An example of this type of confirmatory research is provided within this thesis with the example of the Bodart et..al. (1998) replication.

> *"In the MIS research process, instrument validation should precede other core empirical validities (Cook and Campbell, 1979) which are set forth according to the kinds of questions they answer .. Researchers and those who will utilize confirmatory research findings first need to demonstrate that developed instruments are measuring what they are supposed to measure."*

The term "instrument validity" as defined by Straub relates to the validation of the items and tasks used in the research instrument. A discussion of the development of the instrument will precede the discussion of the validities noted by Cook and Campbell (1979).

Instrument Validation Preceding Study

Straub (1989) makes several recommendations for procedures in developing a valid research instrument. The first recommendation is to look for an obvious reference discipline that is related to the research question of interest with a long-standing tradition of empirical research. The reference discipline chosen for the studies in this thesis is Educational Psychology. This choice is natural for two reasons: 1) the concepts of "understanding" and "learning" used in this thesis are core constructs in Educational Psychology with long standing tradition of research, and 2) instruments for measuring these constructs have been developed previously in this discipline, most notably by Mayer (1989). The use of previously developed instruments improves the potential for instrument validation.

The instrument adopted in this study combines three tasks: comprehension, problem solving, and text reconstruction (Cloze). Individually, these three tasks have been established as reliable measures in Educational Psychology and used in many other empirical studies. The comprehension test has been established as a reliable measure of knowledge and has been used in previous system analysis research studies by Yadav et.al. (1988), Jarvenpaa & Machesky (1989), and Batra, Hoffer, and Bostrom (1990). The problem solving task was developed by Mayer. Mayer (1989, p. 43) reported in his review *of "20 studies involving 31 separate tests*" that the combination of comprehension and problem solving tests "*consistently indicated that models can help lower aptitude learners to think systematically.*" This suggests that the problem solving and comprehension combination has proved to be a reliable measure of what Mayer terms "understanding". In regards to the Cloze test, McKenna & Robinson (1980, p. 12) note in their annotated bibliography that "*Two generalizations have emerged rather convincingly: 1) Cloze does, measure comprehension, and 2) established criteria are relatively stable across populations.*" In terms of the individual tasks, the instrument described in this thesis has been based on a set of measures that have provided reliable results in previous studies.

Choosing valid and reliable instruments does not, however, guarantee a valid instrument. Each of the tasks used in this study had to be adapted to the area of system analysis techniques. The descriptive materials (text descriptions and diagrams) had to be generated, questions had to be created for each of the instruments, and the ability of the instrument to discriminate between treatment groups had to be determined. These three adaptations have been discussed previously in Chapter 5, and are summarized briefly below.

The case materials (diagrams and text descriptions) were created first. Four cases were developed. Three of the cases (IFIP, VOYAGER, and FAR EAST) came from external sources, and the researcher created one case. The text descriptions for the cases were developed first and an attempt was made to keep the total number of words as close as possible to text that would fit on a single page (approximately 350 words). After the text descriptions were created, the researcher and two other persons with experience in system analysis techniques independently created diagrams from the text descriptions[16]. The researcher collected the diagrams and then summarized the drawings from the three sources into a single diagram. The single diagram was then shown the two other persons and changes were made to the diagram as suggested. This iterative process continued until all three parties agreed that the diagrams were a good representation of the text description.

After establishing the descriptive materials, questions were created for the comprehension and problem solving task. No comprehension or problem solving questions had been previously developed for the cases[17]. The researcher developed a set of questions for each case. These questions were then informally tested by a small group of Ph.D. students, who were familiar with questionnaire development, before they were tested on participants[18]. The questions were discussed with this small group in an open interview environment. Suggestions for improvement were incorporated into the instrument. The open interview provided the opportunity for technical validation using a method different from the method that the pilot study participants would use. These questions were then built into the testing program

[16] My thanks again to William Tan and Darrel Jung for their help in developing the diagrams for the cases.
[17] Six comprehension questions from the VOYAGER case noted in Bodart & Weber (1996) were reused.
[18] My thanks to all of the CHAPS, Paul Chwelos, Errol Smythe, and Mike Brydon for their candid comments and suggestions.

described in Chapter 5. A pilot study was then conducted on the instrument, as suggested by Straub (1989).

The participants in the pilot study were paid $10 to complete the tasks. The study was conducted one participant at a time. Early participants were timed and paid an additional $5 for a personal interview of the feedback on the instrument. The personal interview provided a further opportunity for technical validation using a method different from the computer collection method. Ambiguities in questions, spelling and grammatical mistakes, and confusing instructions were eliminated. The number of questions in the comprehension section of the task was also reduced due to the amount of time individuals spent on the task. After modifying the instrument, a larger pilot study of 91 participants was undertaken with the participants roughly split between the two cases: the IFIP and EVENT cases.

The pilot study provided an opportunity to assess the ability of the instrument to discriminate between treatment groups. This ability to discriminate provided an initial indication of the construct validity of the instrument. While the sample sizes were relatively small (less than 16 in each of the treatment groups) differences between treatment groups across both cases were identified, particularly for problem solving and Cloze instruments. These results were also in the anticipated direction. Based on the results from the pre test, two major changes were made. First, participants would complete two cases as opposed to a single case. This provided more opportunity to assess the internal validity of the instrument. Second, participants would be run in groups of up to eight, instead of one at a time, to speed data collection. Independence between the participants was still maintained as each group of participants was randomized, and the lab was monitored by the researcher during the experiment.

The procedures and pilot study noted above and undertaken prior to the collection of the actual study results, help to strengthen the feeling, on the part of the researcher, that inferences can be made regarding the findings developed in the studies reported in this thesis. As Straub (1989, p. 162) notes:

> *Instrument validation is a prior and primary process in confirmatory empirical research...It is important for MIS researchers to recognize that valid statistical conclusions by no means ensure that causal relationships between variables exist. It is also important to realize that, in spite of the need to warranty internal validity, this validation does not test whether the research instrument is testing what the researcher intended to measure. Measurement problems in MIS can only be resolved through instrument validation.*

Instrument validation can never be proved, only disproved. The information provided above yields no obvious indication that the instrument was invalid before the study proceeded. The following sections will discuss the validity of the instrument used during the study.

Construct Validity

Construct validity is the concern that the research constructs underlying the study are accurately reflected in the operations used to collect data about the constructs. The most important example in this study is the following; Do the operationalized measurements "comprehension score", "problem solving score", and "text reconstruction score" used in this study accurately reflect the underlying construct of "understanding?" Much of the weight of this connection falls on the work of Mayer (1989) who argues that the problem solving score, in particular, provides a more "accurate" operation than previously used instruments of understanding which include comprehension and memory recall. Mayer (1989, p. 59) argues:

> *Had we focused on traditional measures such as overall amount recall or overall amount correct on a comprehension test, we would not have found strong differences between model and control groups. What is wrong with overall recall or comprehension performance? These measures are not useful for the present review because the do not provide information concerning how models help students to select, organize, and use scientific information. In contrast, to examine student's* ***understanding*** *requires a focus on the three dependent measures used in this study as well as more fine grained analyses that should be part of future research.*

This statement suggests that, while the three operationalized measures of understanding cannot accurately reflect understanding, these measures are better than the methods used previously. Further, the operationalized measures used in this study are the measures that best reflect the construct of understanding from those currently available. Further work is required to develop more sensitive measures.

Cook & Campbell (1979) identify a variety of potential threats to construct validity. These threats are handled individually in the discussion below. The first of these threats is the inadequate preoperational explication of constructs which is a general problem in system analysis research. The lack of underlying theoretical constructs that serve as a basis for empirical research is the primary problem facing researchers in system analysis. The BWW constructs (Wand & Weber, 1993) may offer a theoretical foundation upon which these constructs can be built. With few underlying theoretical constructs, the lack of well developed empirical constructs reduces the ability of researchers to claim construct validity. The process of instrument validation described in the section above provides some evidence for the operational explication of constructs in this study, but more standardization of these constructs would be useful for future research in analysis methods.

Another potential threat to construct validity is the threat associated with mono-operation and mono-method bias. These threats suggest that using more than a single operation to measure a single construct or using several empirical methods to measure a construct will provide the opportunity to "triangulate" and assess the validity of the operational constructs. For example, the use of three measures of "understanding" is an example of using several operations (comprehension, problem solving, and Cloze) to measure a single construct (understanding). The collection of both protocols data and answers from the three tasks in the intergrammar study is an example of two collection methods. The combination of collection methods also enables the "triangulation" of information from more than one source.

Straub (1989) describes a technical method for assessing construct validity called the multi-trait, multi-method (MTMM) analysis. In an MTMM analysis, the researcher is interested in determining if the measures made across participants are similar across different collection methods. Since this study collected three operations for dependent measures, and used two cases, an MTMM analysis can be developed. What is desired by researchers is a set of operations that are independent of other measures used in the study, and that consistently measure a construct of interest. This ensures that a researcher is not "overlapping" constructs or using too many measures. The test for the MTMM is a simple correlation matrix that includes at least two separate data collection methods (for example two cases). This correlation matrix is shown below in Table 49.

The results of the MTMM analysis in this study are twofold. First, all three measures chosen for this study (comprehension, problem solving, and Cloze) are relatively independent from each other within and across cases. This is shown by the low correlation values that are found in the "off" diagonal terms. These low correlations indicate that the three measures **diverge** from each other and do not measure exactly the same construct. So this finding validates the use of the three measures.

The second finding is that when the measures are applied across the two cases, there is a high correlation between the **same** measure used on **different** cases. This indicates that the empirical instrument is **consistent** across cases. The results in Table 49 below show divergence between the three main measures (low correlations between measures) and convergence within each measures (high correlation between cases.

Table 48: Multi-Trait, Multi Method Analysis

	Case: Event			Case: IFIP		
Variable	EVCMPTOT (Case 1)	EVPRBTOT (Case 1)	EVCLZTOT (Case 1)	IFCMPTOT (Case 2)	IFPRBTOB (Case 2)	IFCLZTOT (Case 2)
EVCMPTOT (case 1)	1.00 p=.000	X	X	X	X	X
EVPRBTOT (case 1)	-.099 p=.311	1.00 p=.000	X	X	X	X
EVCLZTOT (case 1)	.084 p=.392	.217 p=.025	1.00 p=.000	X	X	X
IFCMPTOT (case 2)	**.268 p=.006**	.001 p=.989	.191 p=.049	1.00 p=.000	X	X
IFPRBTOB (case 2)	.136 p=.171	**.626 p=.000**	-.022 p=..817	.095 p=.329	1.00 p=.000	X
IFCLZTOT (case 2)	.047 p=.628	.147 p=.132	**.649 p=.000**	.129 p=.188	-.056 p=.565	1.00 p=.000

In this study, the MTMM analysis provides evidence that the three dependent measures are divergent, and that each of the three operations are consistent measures.

Other threats to construct validity include confounds, interaction of different treatment groups, and interaction of testing and treatment groups. These were reduced by the set of experiment procedures described in Chapter 5. In general, the care and attention in creating the instrument and observing the experimental procedure gives this researcher reasonable confidence in the construct validity associated with this study.

Reliability

Reliability is an important consideration regarding the operationalized measures used in an experimental study. Participants were required to complete two cases instead of a single case as an additional assessment of the reliability of the measures used in the study. Since two cases were used for each study, it is possible to observe the relative measures across cases used in the same study. An example is useful here. Two cases are used in the OOA and DFD/ERD comparison in this study. The Hypothesis for this study states that OOA participants should perform better than DFD/ERD participants on the problem solving task. Three outcomes could potentially be observed. First no difference between OOA ad DFD/ERD may be found in both cases. If we assume that the theory regarding the difference is true, then the finding of "no difference" indicates a problem of validity in the instrument. In other words, the instrument is not measuring what it intends to measure. If, on the other hand, a difference is found in one case but not the second case, then this indicates a problem with reliability. The instrument does not consistently provide the results. In the third case, differences could be observed in both cases. When differences are consistently observed, validity cannot be established, but the instrument has not been proven to be unreliable. The results from both of the studies in this thesis are consistent. There is no indication of statistically significant differences in the results across cases. These finding suggest that the instrument developed in this thesis has provided consistent results and has not been proven unreliable.

The argument for reliability can be taken one step further. The replication of the results in the intragrammar study by Bodart et.al. (1998) show the same consistent results as described in this thesis. Bodart et.al. (1998) extended the results to three studies, five different cases, and two separate research teams. In all cases the results indicated the same predicted direction and relative difference. Reliability is difficult to establish, but the consistent results observed in multiple cases across different research teams, using measurement instruments that have an established tradition of reliability, provides some strong initial evidence for the reliability of the instrument suggested in this thesis.

Internal Validity

Internal validity is threatened when the observed outcomes in a study can be inferred from a variety of causal sources. In other words, there are other explanations for the outcomes than those proposed in the study. Common examples of these threats such as maturation, mortality, diffusion of results, and participant history are not relevant to this study. Instrumentation is another threat that can be excluded due to a single measurement. One important threat in this study is the effect of case order on results. A previous analysis in Chapter 6 showed there was no significant difference in any of the dependent variables within treatment groups across different orderings of the cases.

Another possible threat is the selection bias or non-response bias associated with participants. In this case, participants were taken from a pool of 3rd and 4th year students who had at least one MIS course. The students were not required to participate and were paid $15 for each student participating. A further incentive of $25 for the top 8 scores in the study was offered as motivation for a good effort. A large percentage (over 60%) of the students who were approached to take the study chose to take the study. No systematic bias was apparent in the individuals taking the study. An analysis of the demographics of the participants showed a wide variety of participants and no systematic bias between groups. The analysis of demographics was reported earlier in section 6.3, and 6.4 in Chapter 6.

Evaluation apprehension is another threat to internal validity. In this threat, participants answer what is "expected" by the researcher rather than their true response. The researcher in not divulging expected results controls this. This threat was minimized in this study by reducing the interaction between the researcher and the participant and eliminating the interaction between participants. Since participants were not aware of the differences in the treatment, it was not obvious what results the researcher wanted to observe.

Another potential bias affecting internal validity is experimenter's expectancies. This is particularly important in coding the problem solving scores or the coding of protocol analysis as these are the most subjective dependent measures. The experimenter's expectancies were reduced by creating a "blind" coding environment with no obvious indication of what treatment groups a participants belonged to. Unfortunately, it was impossible to maintain the "blind" environment during the collection of protocols, and a potential for experimental expectancies is recognized. To reduce the level of subjectivity with the coding of methods, a second independent person, MBA student, coded all of the protocol, responses.

A high degree of reliability was observed in the two studies reported in this thesis. The high inter-rater reliability observed in the coding of the problem solving task indicated that the problem solving measures provided consistent results. The Pearson correlation coefficient of 0.90 and 0.89 was observed for the two cases used in the intergrammar study, and 0.92 and 0.90 for the cases in the intragrammar study. To reduce the threat of experimenter expectancy even further, the coding created by the independent raters was used in the study instead of the coding used by the researcher.

The experimental controls maintained independence between participants. The randomization of participants across treatment groups, the randomized order of cases, the blind coding of the problem solving scores, and other procedures described above give the researcher confidence that the intergrammar and intragrammar studies reported earlier in Chapter 5, 6, and 7 have no obvious indications of problems with internal invalidity.

Statistical Conclusion Validity

Statistical conclusion validity is established when the operationalized variables demonstrate relationships that are not explained simply by chance events. This type of validity addresses whether there are significant differences between treatment groups. The most common threats to statistical conclusion validity include violations in the assumptions underlying statistical procedures, low statistical power, low reliability of measures, and low reliability of treatment measures. These threats have been discussed in length in Chapters 5, 6, and 7 and the discussion is summarized below.

The MANCOVA procedure was introduced in chapter 6 along with a discussion of the assumptions underlying this procedure. The six assumptions underlying MANCOVA and the tests for the violations of these assumptions were summarized in Figure 10 in Chapter 7. In general, the assumptions were not violated except for the assumption of homogeneous covariance matrices. As discussed in Section 6.1, the large sample size and relatively equal group sizes in this study reduce the impact of the violation of this assumption. The effect of the violation is to decrease the power of the test. Type 1 error rates are affected only slightly except in extreme cases, and there is nothing to indicate that the conditions in this case are extreme.

The power of the tests reported in the study is another important consideration. The power of the test is the probability of accepting a null hypothesis that is false. This estimate is related to sample size with larger sample sizes, all other things equal, providing larger power. Power is also inversely related to alpha (α), the level of type I error, so as alpha is made smaller, the level of type II error (β) becomes larger and power ($1 - \beta$) is decreased. Power is also affected by the size of the effect, which is the distance of the observed mean from the hypothesized mean. Larger effects will have larger power as we are more likely to detect large rather than small discrepancies. While the level of acceptable type II error rates is a matter of preference, Cohen (1969) indicates that, in general, power levels below 80 percent are inconclusive and do not indicate the effect is present. Baroudi & Orlikowski (1989) also suggest the 80 percent level in their study of MIS research. In this study, the power levels associated with the multivariate effects are very high, in the 0.990 range. The power associated with univariate statistics are naturally lower ranging between 0.69 and 0.89 for an $\alpha = 0.05$, for all significant effects, depending on the size of the effect. The power level in this study are generally acceptable, however, an increased sample size for each treatment groups would increase the power estimates for univariate tests.

Poor reliability might also threaten statistical conclusion validity. Reliability has been discussed earlier in this section. Since none of the covariates were used in the final estimates, the reliability associated with the scale variables, (knowledge of method, knowledge of domain) would not effect the conclusions. The reliability of the problem solving scores was shown to be high, as the correlation between the two independent raters was over 0.90 which is significant at $\alpha = 0.0001$. The degree of agreement between results reported across cases and the confirmation of results from Bodart et. al (1998) provide the researcher with a reasonable level of confidence in the reliability of the measures and treatment implementation.

This section has focused on the results from the MANOVA analyses. The result from the protocol analysis should also be examined. Less statistical conclusion ability can be attributed to the results of the exploratory protocol analysis for several reasons. These reasons include the lack of verified measures and prior studies with a clearly identified unit of analysis, and the subjectivity associated with the analysis of these findings. The use of an independent rater for some of the more "objective" measures provides more confidence in these findings. While the researcher is somewhat less comfortable with the validity of the results from the protocol analysis, the researcher finds the results from the protocol analysis compelling. Further studies and experience with this method will improve the level of confidence and view of validity associated with this methods.

In summary, the general convergence of the tests to the assumptions underlying MANOVA and ANOVA procedures, the comfortable levels of both α and β used and observed in the test, and the reliability of the results across cases that were observed under controlled experimental conditions provides the researcher with reasonable confidence in the statistical validity of the results reported in the intergrammar and intragrammar comparisons.

Summary

This section has outlined a sequence of arguments regarding the validation of the empirical instrument developed in Chapters 4 and 5 of this thesis. The discussion of validity began with the development of the instrument and the procedures used to improve the validity of the instrument before the test. These steps included using previously validated instruments related to the problem solving, comprehension and Cloze tasks, developing test materials with several individuals knowledgeable in systems analysis, and implementing a pilot study of the instrument with 91 individuals. Further refinements of the instrument as a result of the pilot test and results from the pilot test indicated a reasonable validation process preceding the study.

The work by Mayer (1989) along with the design of the experimental controls was used as a basis for forming the arguments for construct validity. The existence of a similar set of instruments, along with a model of the learning process, and an argument for why problem solving scores would differ between groups enables the researcher in this study to build on the foundation supplied by Mayer. The use of several tasks (comprehension, problem solving, and Cloze), several cases, and even different methods (protocol analysis) improved the argument for construct validity. This argument for construct validity is closely related to the issue of reliability that was addressed above. In general the reliability of the measures was shown with a single measures (for example the high inter-rater reliability for the problem solving score), between cases, and even between research groups.

Having made a case for a reasonable level of construct validity and reliability in the measures used in the study, the internal validity of the two studies was discussed. Internal validity is the issue of whether there exist reasonable alternative hypotheses for the observed effects. The issue of internal validation is largely addressed in the experimental design, where an attempt is made to reduce or eliminate extraneous effects. Chapter 5 discusses experimental design in detail. In general, the independence between participants, randomization of participants into treatment groups, large sample sizes, blind coding of problem solving, and control over materials and collection methods enable the researcher to feel a reasonable level of confidence regarding the internal validity of both the intergrammar and intragrammar study.

Finally, after considering the internal validity of the study, the statistical conclusion validity of the results obtained from the study was considered. The testing of Hypothesis, the generally satisfactory power of test associated with statistically significant findings, and the reliability of findings as observed across cases and research teams (Bodart et.al., 1998) provide a comfortable level of confidence in the statistical results produced by the study.

Together, the experiment and discussion of validation preceding the study, construct validity, reliability, internal validity, and statistical conclusion validity discussed above have helped to establish a reasonable argument for the validity of the empirical instrument described in Chapters 4 and 5 of this thesis. These results enable us to make the following conclusion regarding the empirical instrument that is at the heart of this study:

1. ***The procedures described and the data observed in the two studies, provides evidence that the empirical instrument developed in Chapter 4 and 5 of this thesis is capable of differentiating the level of "understanding" (as measured by the combination of comprehension, problem solving, and Cloze scores) for both intergrammar and intragrammar comparisons. These results, in conjunction with the validity of the instrument and experimental procedures as discussed above, provides evidence that the empirical instrument has a acceptable degree of validity and reliability in the context of this study.***

The motivation of the thesis was to develop a set of empirical techniques that could be used to compare system analysis methods. The instrument described in this study has been shown to provide the necessary sensitivity to make these types of comparisons possible. Even further, the instrument has been argued to possess a reasonable level of validity. While there are still challenges ahead in developing standard instruments for measuring domain knowledge, analysis method knowledge, and ease of use, the instrument proposed in this thesis has satisfied this researcher's initial objective.

8.2 Results Drawn from the Studies

Having discussed the instrument used in the two studies and having addressed the issues of validity related to the instrument, we turn to the second objective to apply the comparative empirical instrument discussed above to evaluate actual grammars. Two studies were undertaken. The first study, an intergrammar comparison, drew a comparison between text description (TXT), Object Oriented Analysis (OOA) and a combination of Data flow Diagramming (DFD) and Entity Relationship Diagramming (ERD) grammars. The second study, an intragrammar study, revisited the study previously reported by Bodart & Weber (1996) to compare "Mandatory" and "Optional" grammars of the Entity Relationship Diagram

(ERD). The studies designed, implemented, and analyzed in Chapters 5, 6, and 7 of this thesis enable the researcher to make the following conclusions regarding the initial hypotheses made in Chapter 4. Conclusions numbered 2-6 below rely on the initial result regarding the validity of the instrument noted in (1) above.

Summary of Results

2. ***The data collected in this study provides evidence supporting the hypothesis that graphical diagrams communicate at least the same amount of comprehensive knowledge, and lead observers to a potentially higher level of understanding regarding a domain, in the same amount of time as text descriptions. The data also suggest that the level of understanding gained by individuals viewing a graphical model, vary with the diagram being used.***

3. ***The data collected in this study provides evidence supporting the hypothesis that participants viewing Object Oriented Diagrams score significantly higher in problem solving scores, and NOT significantly different in Cloze or comprehension tests when compared to participants viewing the combination of a Data Flow Diagram and a Entity Relationship Diagram. Further, the participants using the Object Oriented Diagram do not take significantly different time to complete the study tasks when compared to participants provided with a combination of Data Flow Diagram and Entity Relationship Diagram.***

4. ***The observations in the protocol analysis indicate that for even simple problems the connection between DFD and ERD is difficult for individuals to make. Further, novices tend to gravitate towards "object-like" constructs when reasoning about a system. And when provided with an OOA model, participants tend to organize and refer to the OOA more often than DFD/ERD participants refer to the DFD or ERD. Together these observations suggest a possibility that that the object model may be a more "natural" approach to system analysis than the separation of data flows from data structure recommended by structured analysis. This is only one possible explanation, and further research needs to be done to understand the nature of this advantage. The finding do suggest that object oriented analysis has the potential, to improve how individuals reason about and come to an understanding of a domain.***

5. ***The data collected in this intragrammar study provides evidence supporting the conclusion that ER diagrams created with the "Mandatory" grammar communicate at least the same amount of comprehensive knowledge as diagrams created with an "Optional Grammar. Further, the participants viewing diagrams created using the "Mandatory" grammar score a measurably higher level of "understanding", as measured in problem solving and Cloze scores, in the same amount of time, as diagrams created with the "Optional" grammar.***

6. ***The data collected in the intergrammar and intragrammar studies provide evidence supporting the hypothesis that neither the observed level of prior domain knowledge nor the observed level of analysis method knowledge had a significant effect on the level of "understanding" that participant's developed regarding a domain. This***

conclusion should be viewed with skepticism, however, due to the lack of valid and verified measures for domain and analysis method knowledge.

These propositions were expressed earlier in chapter 7 and have, hopefully, gained more validity in light of the discussion of the validity of the instrument used to collect these results. In the next section the implications of these findings are discussed.

8.3 Implication of the Research Findings

The implications discussed in this section are driven by the original motivation for the thesis. The motivation was to argue for increased attention focused on the comparison of analysis techniques in an effort to reduce the tendency towards method proliferation and to improve the performance of analysis techniques. It was argued that rather than concentrating effort on producing a wider variety of analysis techniques, researchers should attempt to understand how and why analysis techniques are useful in the design process. This understanding may help to more completely define the roles that analysis techniques play in information system development, and potentially help to produce more effective analysis techniques.

Implications Regarding Empirical Comparisons

1. The success of the instrument developed in this thesis and discussed in this chapter implies that future empirical comparisons of analysis techniques must go beyond the traditional use of comprehension tests and move towards methods that supply information about a "higher" level of understanding. The success of the problem solving measure and the exploratory protocol analysis reported in Chapter 7, are examples of the types of measures capable of capturing glimpses of the "higher" level. Note that no difference between the OOA and DFD/ERD combination would have been observed without this information. Mayer's (1989) methods present a new technique that should be applied in future comparisons.

2. Previous empirical studies such as Jarvenpaa & Machesky (1992) and Vessey & Conger (1994) have indicated the need for multiple measures when developing comparisons. The results obtained in this study confirm that multiple measures (for example problem solving, comprehension, Cloze) are necessary to gather a more complete notion of the differences.

This idea should be extended to the collection using multiple methods such as protocol data. The multiple methods not only provide the ability to validate measures through triangulation, the methods can also reveal information not available from questions and tasks that have been used in previous research.

Implications Arising from Intergrammar Comparison

3. The comparisons between text descriptions and graphical models (diagrams) suggest that some graphical models facilitate a "higher" level of understanding regarding a domain than text descriptions. This implies that diagrams are useful means of communicating domain information to individuals, regardless of whether they understand the grammar underlying a diagram or not. The fact that the comparison between object-oriented analysis (OOA) and text descriptions showed significant differences while comparisons between the combination of data flow diagram (DFD) and entity relationship diagram (ERD) with Text descriptions did NOT show significant differences implies that not all diagrams are created equal, and not all diagrams, therefore, may deliver more conceptual knowledge than a text description.

4. The comparison between OOA and DFD/ERD suggests that OOA diagrams facilitate a "higher" level of understanding regarding a domain over the combined DFD and ERD. This implies that the grammar underlying the OOA diagrams possesses certain advantages over the two grammars in the DFD/ERD combinations. This result provides the first empirical evidence (to the best of the researcher's knowledge) of the advantage of an OOA grammar over more the more traditional DFD/ERD combination.

5. The analysis of the protocols from DFD/ERD and OOA participants indicates that the source of the advantage of the OOA over the DFD/ERD combination is evident in two areas. First the DFD/ERD participants made very little use of the ERD and hence may have lacked a structural component in their thinking. Second, the OOA participants made more reference to the OOA diagram, and were observed to use the OOA diagram as a conceptual framework more often than the DFD/ERD participants made use of the DFD. This implies that the OOA constructs may be a more "natural" conceptual construct than the constructs in the DFD. An analysis of the diagrams created by participants tended to support this claim. These results, in general, imply that the object-oriented grammar possesses advantages over the traditional DFD/ERD combination. Object–oriented

analysis techniques should receive additional attention from both researchers and practitioners.

Implications Arising from the Intragrammar Study

6. The comparison between "Mandatory" and "Optional" grammars for the Entity Relationship Diagram (ERD) suggests than the "Mandatory" grammar better facilitates the development of a "higher" understanding of a domain than the "Optional" grammar. These results have been confirmed in a study by Bodart et.al. (1998). This result implies that the "Mandatory" grammar should receive more attention from researchers and practitioners, and that "Mandatory" grammars should be recommended for those choosing to developing schema in ERD.

7. The theoretical difference between the "Mandatory" and "Optional" grammar is provided by the argument based on ontological constructs described in the Bunge-Wand-Weber (BWW) ontology. The fact that differences were observed in the direction predicted by the BWW ontology provides confirmatory evidence that the ontological constructs are capable of being used to predict empirical differences. While this evidence cannot confirm the BWW ontology, the test lends more credibility to the BWW constructs.

8.4 Contributions of the Study

The thesis has contributed to the study of system analysis methods in several ways. First, the empirical instrument described in this thesis extends the previous empirical research instruments with the introduction of the problem solving task form Mayer (1989) and the Cloze task from Taylor (1953). In the two studies described in this thesis, the instrument has displayed an improved ability to differentiate between treatment groups over the more traditionally used comprehension test. The success of the instrument in differentiating between treatment groups suggests that future empirical comparisons in system analysis should consider using these techniques as a part of the comparison.

The second contribution is the first empirical evidence (to the best of the researcher's knowledge) that object-oriented analysis (OOA) grammars have an advantage over more traditional "structured analysis" combination of the Data Flow Diagram (DFD) and the Entity

Relationship Diagram (ERD) in the process of interpretation from a diagram. The exploratory protocol study provided some indications why these advantages were observed. These indications include: the observation that participants were not able to easily make the connection between the DFD and the ERD; that participants using OOA were more likely to structure their answers using the OOA as a guide than the DFD/ERD participants; and that participants drawing their own diagrams of the domain tended to use “object-like” structures in their diagram regardless of what treatment group they were in. These results confirm the arguments made by proponents of object-oriented methods that objects are more “natural” constructs that the previous ‘Structured analysis” constructs.

The third contribution made by this study was the result that individuals viewing ERD's created with attributes that have "Mandatory" properties and entity subtypes outperform individuals provided with ERD's created with optional properties in the interpretation process. This result provided some confirmatory evidence for the argument that the Bunge-Wand-Weber (BWW) ontology (Wand & Weber, 1993) has predictive capabilities. This result extended previous research in the field of system analysis, Bodart & Weber (1996). In a later study, Bodart et. al. (1998) confirmed the results found in this thesis, providing a rare opportunity to independently confirm the findings presented in this thesis, before the thesis was presented.

Other, somewhat less important contributions include the confirmation of the common sense notion that graphical models out perform text descriptions in interpretation, and that all graphical techniques are not created equal. The study also indicated that for the cases presented in the studies, our measures of the previous knowledge of system analysis techniques and prior knowledge of a domain were not significant factors in explaining the level of comprehension, problem solving, or the Cloze results. This observation is a reflection of the lack of a good set of empirical measures for domain and modeling knowledge, or perhaps the relative ease of the cases,

8.5 Limitations of the Study

The discussion of the limitations of the study will focus on the question of external validity of the study results. External validity is the extent to which the results of this research can be generalized to different participants and experimental settings. When conducting an experiment, a natural tradeoff between external validity and internal validity arises. Internal validity seeks to restrict the experimental environment to reduce the number of extraneous effects on the observed results. The restrictions on the environment consequently leads to an environment that does not resemble "real world" conditions. This limits the level of external validity that can be associated with the results. Some of the restrictions placed on this study are noted below.

The first restriction is that the participants used in this study were not experienced system analysts. The study used students who either had one course in systems analysis or no course at all. The study results should not, therefore, be extended to experts in the system analysis field. Since interpretation was the primary process being measured, this restriction need not diminish the impact of the research results, as many of the persons viewing the analysis diagrams will be novices (system users and manager) rather than experts in analysis techniques.

The second restriction imposed on the study was to limit the complexity of the cases. The cases were designed to be small enough to understand in a time reasonable for the study. The study results should not be extended, therefore, to real world problems that are extremely complex. This is an obvious limitation of the study, but one that is necessary for the logistics of the experimental procedure. This does not, however, completely discount the results of the study. It could be argued, for example, that if the differences in performance observed in this study were apparent for even simple cases, there is at least a prima facie case that the differences may be magnified for more complex cases. The complexity of the case is a difficult measure to make, and arguments regarding the level of complexity are difficult to resolve. The researcher's intention is to begin with simple cases, to compare techniques at a basic level, and to use this information to build towards more complex cases. The fact that the cases are not as complex as those in the real world, does not discount the differences observed, but might limit the extension of the results to more complex situations.

A third consideration is the grammars used in the study. It is important to note that the comparisons are based on ***specific*** grammars and that care should be taken in extending the results of the comparisons of a specific grammar to all related grammars. For example, the object-oriented grammar used in this study was found to perform better than the combination of data flow diagrams and entity relationship diagrams. This does not imply however, that ***all*** object-oriented grammars outperform ***all*** grammars based on the DFD/ERD combination. Some indication has been given regarding the advantages of object–model constructs, but these results should not be generalized without more confirmatory evidence.

These three considerations are the largest limitations placed on the study results. In general, the external validity of the results should be limited to only those cases that share the same level of complexity, and where novices are the primary interest of the study.

8.6 Directions for Future Research

The small number of empirical comparisons of system analysis experiments discussed at the start of this thesis must be addressed by future researchers. The lack of standard measures and procedures makes the development and implementation of empirical comparisons of analysis techniques a challenge. This thesis has shown that if an empirical instrument can be created and established, as this thesis purports to do, there are a wide range of potential comparisons that can be made. Some of the potential comparisons are discussed below.

1. The Unified Modeling Language (UML) has been proposed as a standard language for the approach to system analysis (Booch, Jacobsen, and Rumbaugh, 1997). The UML is increasingly popular, yet very few empirical research studies have been undertaken to verify the utility of the approach (Castelani, 1998). The wide variety of grammars associated with the numerous diagramming techniques suggests that the UML may be difficult for users to understand, in a way similar to the DFD/ERD combination. The instrument proposed in this study, could be used to evaluate the interpretation from UML diagrams in the same way that the OOA and DFD/ERD combination were compared in this study.

2. The participants used in this study were students with a varying degree of experience with analysis methods. A logical step would be to move towards a comparison based on professional system analysts with experience in actual projects. This was suggested by Jarvenpaa & Machesky (1989) and Vessey & Conger (1994). While there are obvious biases associated with expert participants, the results would provide a view of how the analysis techniques would perform for practitioners.

3. This study has focused almost exclusively on the interpretation from a diagram that has already been created. A full evaluation of an analysis technique should also include an analysis of the representation of a diagram. For example, if a diagram is easy to interpret but takes a very long time to create when compared to other techniques, the analysis

technique might not be particularly useful for practitioners. The difficulty in extending the study towards the process of representation is the need to again develop appropriate measures, and also the need to use "expert" analysts as the results from novices would likely have very little external validity. Since the process of interpretation is generally left to system analysis "novices" it is natural to focus on novices for interpretation. But very few novices will be asked to create representations. For this reason, the use of experts is required for the representation process. Of course experts also come with a well defined bias towards methods they are used to, and this bias must be overcome for new methods. The smaller number of experts also makes this type of study a difficult challenge.

4. System analysis methods and techniques are primarily "static" models created on pen and paper or on a computer screen. Researchers have yet to suggest, to the best of this researcher's knowledge, the use of animation and narration associated with system analysis techniques. The extension from "static" paper based diagrams to "active" animated models with narration seems a natural step to make for system analysis techniques. Current technology would support the development of these methods, and the instrument developed in this thesis could be used to evaluate the efficacy of using animation and narration in system analysis techniques. Mayer and Anderson (1991,1992) have demonstrated the improved product of animation in describing explanative information in the field of educational psychology. Mayer & Anderson (1991,1992) further suggest that animated models combined with narration provide a higher level of "understanding" to a participant than either narratives or animations alone. They suggest that the improved product is the result of a ***contiguity principle***, which states that " *the product of multimedia instruction increases when words and pictures are presented contiguously (rather than isolated from one another) in space and time."* More importantly, Mayer & Anderson (1991,1992) show that the animations and narratives lead to superior scores in tests for understanding when compared to static, paper based descriptions of domains.

It is recognized that further development of the empirical instrument described in this thesis is required in order to improve the validity and reliability of the measure. These adjustments, however, represent minor rather than major updates to the procedure. The suggestions listed above indicate a significant potential for further studies in system analysis that would result from the development of an acceptable instrument.

8.7 Summary

This chapter began by stating two objectives. The first objective was to argue for the validity of the empirical instrument described in this study. The discussion in section 8.1 and 8.2 provided evidence for the conclusion that the instrument and related procedures used in this thesis can be considered valid and reliable. This is the primary objective of the thesis. It is left to the reader to determine whether this has been argued successfully.

The second objective was to use the empirical instrument to perform some interesting comparisons between analysis methods. Two studies were implemented, and the observations drawn from the study both confirm, in general, the proposed differences and indicate that the empirical instrument is sensitive enough to detect these differences. Implications were then drawn from these observations. Primary among these were the implication that object-oriented methods may possess advantages, in certain domains, over the traditional structured analysis approach, and that a grammar featuring mandatory attributes and relationships with subtypes is preferred over the a grammar using optional attributes and relationships in the entity relationship model. A set of limitations to the study was also discussed.

The chapter then ended with a discussion of future research directions. This short discussion revealed a large area of research that could be opened with the development of a valid instrument for comparing analysis methods. While further refinements in the empirical instrument used in this thesis are important, the utility of the instrument seems to be evident, at least in the somewhat biased view of this researcher, in the studies described earlier.

As a final note, it is my hope that the instrument and method described within will prove useful enough to provide information for improving system analysis methods, and more importantly, to improve the communication between system users and system developers which is a necessary ingredient in system development success.

Bibliography

Avison, D.E., and G. Fitzgerald, Information Systems Development: Methodologies, Techniques, and Tools, Second Edition, McGraw-Hill, London,1995.

Batra, Dinesh, Jeffrey Hoffer, and Robert Bostrom "Comparing Representations with Relational and EER Models" Comm. of the ACM, Vol. 33, No. 2, Feb. 1990, p. 126-139.

Bodart, Francois, Avind Patel, Marc Sim, and Ron Weber, "Should The Optional-Property Construct Be Used In Conceptual Modelling?: A Theory and Three Empirical Tests." Unpublished Working Paper, The University of Queensland, December 1998.

Bodart, Francois and Ron Weber "Optional Properties Versus Subtyping in Conceptual Modeling: A Theory and Empirical Test", Proceedings International Conference on Information Systems, Dec. 16-18, 1996, pp. 450.

Boehm, B., "Software Engineering", IEEE Transactions on Computers, Vol. C-25, Dec. 1976

Booch, Grady, *Object-Oriented Analysis and Design with Applications*, Rational, Santa Clara, CA., Second Edition, 1995, Chapter 2, pp. 34-38.

Booch, Grady, Ivar Jacobsen, and James Rumbaugh, "Unified Modeling Language (UML), Version 1.1", Rational Software Corporation, Santa Clara, California, http://www.rational.com, accessed October 10, 1997.

Bormuth, John R. "Comparing Cloze and Multiple Choice Comprehension Tests", Journal of Reading, Volume 10, Number 5, February 1967, pp. 291-299.

Brooks, F. P. Jr. "No Silver Bullet: Essence and Accidents of Software Engineering," *Computer 20,* 4 (April 1987): 10-19.

Brosey, Margaret, and Ben Shneiderman, "Two Experimental Comparisons of Relational and Hierarchical Database Models", Inter. Journal of Man-Machine Studies, Vol. 10, 1978. p. 625-37.

Bunge, Mario, *Treatise of Basic Philosophy: Volume 3: Ontology I: The Furniture of the World*, Reidel, Boston, 1977.

Bunge, Mario, *Treatise of Basic Philosophy: Volume 4: Ontology II: A World of Systems*, Reidel, Boston, 1979.

Castlini, Xavier, "Evaluation Models Defined with Charts of Concepts: Application to the UML Model", Proceedings of the CAiSE'98 / IFIP 8.1 International Workshop on Evaluation of Modelling Methods in System Analysis and Design, ed. Siau, Keng, Pisa, Italy, June 8-9, 1998, Section E pg. 1-12.

"Chaos", Standish Group Report on Information System Development, 1995, http://www.standishgroup.com/chaos.html [Accessed July 12, 1996].

Chatzoglou, P.D. and L.A. Macaulay, "Requirements Capture and IS Methodologies", Information Systems Journal, 1996, 6, pp. 209-225.

Coad, Peter and Edward Yourdon, *Object Oriented Analysis*, Yourdon Press ,1991

Cook, T.D. and D.T. Campbell, *Quasi-Experimentation: Design and Analysis for Field Settings*, Houghton Mifflin Company, Boston, 1979.

De Marco, Tom. *Structured Analysis and System Specification*, Prentice-Hall Inc., Englewood Cliffs, 1979.

Ericcson K.A., and H.A. Simon, *Protocol Analysis: Verbal Reports As Data*, Revised edition, the MIT Press, 1993.

Floyd, C., “A Comparative Evaluation of Systems Development Methods”, in *Proceeding of the IFIP WG 8.1 Working Conference on Comparative Review of Information Systems Design Methodologies: Improving the Practice.*, Olle, T., W.Sol, H.G. and Verrjin-Stuart, A.A. (eds), North Holland, Amsterdam, 1986.

Gabig, Jerome S. “Software Development Contracting is Risky Business; Buying the Secret Codes”, Legal Times, June 17, 1996, Pg. S42

Gane, Chris and Trish Sarson, “Structured Systems Analysis: Tools and Techniques”, Prentice-Hall, Inc., Englewood Cliffs, 1979.

Gemino, Andrew “To Be or May to Be: An Empirical Comparison of Mandatory and Optional Properties in Conceptual Modelling”, Proceedings of the Administrative Sciences Academic Conference, June 4-6, 1998, Saskatoon, Sask, 31-43.

Gemino, Andrew and Yair Wand, “Empirical Measures for Comparing System Analysis Methods”, Proceedings of the CAiSE‘98 / IFIP 8.1 International Workshop on Evaluation of Modelling Methods in System Analysis and Design, ed. Siau, Keng, Pisa, Italy, June 8-9, 1998, Section H pg. 1-12.

Gemino, Andrew, “To Be or May to Be: An Empirical Comparison of Mandatory and Optional Properties in Conceptual Modelling”, Proceedings of the Annual Conference of the Administrative Sciences Association of Canada, Information Systems Division, Volume 19, No. 4, Saskatoon, Saskatchewan, May 30- June 2, 1998, pg. 33-44

Gemino, Andrew and Yair Wand, “Empirical Comparison of Object-Oriented and Dataflow Models”, Proceedings of the Eighteenth International Conference of Information Systems, eds. Kumar, Kuldeep, DeGross, Janice, Dec. 15-17, 1997, Atlanta, Georgia, Paper no. 32, pg. 446-447.

Gemino, Andrew, “Taming the Methodology Jungle: An Argument for Empirical Research in Analysis and Design Modelling”, accepted as presentation for Annual Conference of the Administrative Sciences Association of Canada Information Systems Division, St. John’s, Newfoundland, May 28-30, 1997.

Gemino, Andrew, “The Empirical Analysis of System Analysis and Design Methodologies”, Electronic Proceedings from International Conference of Information Systems Doctoral Consortium, December 8-10, 1995, Amsterdam, Netherlands.

Gibbs, W. Wayt, "The Software Crisis", Scientific American, September, 1994, p. 86-95.

Gick, Mary and Kieth Holyoak, “Analogical Problem Solving”, Cognitive Psychology, Vol.12, 1980, p. 306-355.

Green, Peter, *An Ontological Analysis of Information Systems Analysis Design (ISAD) Grammars in Upper CASE Tools*, Ph.D. Thesis, Department of Commerce, University of Queensland, Australia, May 1996.

Hare, Joseph, Rolph Anderson, Ronald Tatham, and William Black, *Multivariate Data Analysis*, Third Edition, Macmillan Publishing Company, New York, 1992.

Hirschheim, Rudy, Hienz Klien, and Kalle Lyytinen, “Exploring the Intellectual Structures of Information Systems Development: A Social Action Theoretic Analysis”, *Accounting, Management & Information Technology*, 1996, Vol. 6, no. 1/2, pp. 1 - 64.

Hoffer, George, Joey George, and Joe Valacich, “ Modern System Analysis and Design”, Benjamin/ Cummings Publishing, Don Mills, Ontario, Second Edition, 1998.

Hutchings, A.F. and S.T. Knox, “Creating Products Customers Demand”, Communications of the ACM, Vol. 38, No.5, May 1995, pp. 72-80.

Holdblatz, K., and H.R. Beyer, “Requirement Gathering: The Human Factor”, Communications of the ACM, Vol. 38, No.5, May 1995, pp 31-32.

Jacobson, Ivar, Maria Ericsson and Agneta Jacobson, The Object Advantage: Business Process Reengineering with Object Technology, Chapter 3, pp.46-76.

Jarvenpaa, Sirka and Jeffrey Machesky, "Data Analysis and Learning: An Experimental Study of Data Modeling Tools", Int. Jrnl. of Man-Machine Studies, 31, 1989, pp 367-91.

Kemerer, C.F. “An Empirical Validation of Software Cost Estimation Models”, *Communications of the ACM*, Vol. 30, Number 5, May 1987, pp. 416-429.

Lindland, Odd Ivar, Guttorm Sindre, and Arne Solvberg, “Understanding Quality in Conceptual Modeling”, IEEE Software, March 1994, p. 42-49.

Lucas, Henry C., *Why Information Systems Fail*, Columbia University Press, 1975.

McFarlan, F. W., “A Portfolio Approach to Information Systems”, *Harvard Business Review*, Volume 59, Number 5, September-October, 1981, pp. 142-150.

Mayer, Richard E. "Can you repeat that? Qualitative Effects of Repetition and Advance Organizers on Learning from Science Prose", Journal of Educational Psychology, 45, 1982, pp. 40-49.

Mayer, Richard E., "Models for Understanding", Review of Educational Research, Spring 1989, Vol. 59, No. 1, pp 43-64.

Mayer, Richard E., *Thinking, Problem Solving, Cognition*, Second Edition, W.H. Freeman and Company, New York, 1992.

Mayer, Richard E., "Learners as Information Processors: Legacies and Limitations of Educational Psychology's Second Metaphor", Journal of Educational Psychology, Vol. 31 (3/4), p.151-161.

Mayer, Richard E. and R.B. Anderson, "Animations need Narrations: An Experimental Test of the Dual-Coding Hypothesis", Journal of Educational Psychology, 1991, Vol. 83, no.4, pp. 484-490.

Mayer, Richard E. and R.B. Anderson, "The Instructive Animation: Helping Students Build Connections Between Words and Pictures in Multimedia Learning", Journal of Educational Psychology, 1992, Vol. 84, no.4, pp. 444-452.

Mayer, Richard E., William Boye, Alexandra Bryman, Rebecca Mars, and Lane Topango, "When Less is More: Meaningful Learning from Visual and Verbal Summaries of Science Textbooks", Journal of Educational Psychology, vol. 88, No. 1, 1996, p. 64-73.

Mayer, Richard E. and J.K. Gallini, "When is an Illustration Worth Ten Thousand Words?", Journal of Educational Psychology, 1990, Vol. 82, no.4,pp. 715-726.

McKenna, M.C., and R, Robinson, An Introduction to the Cloze Procedure: An Annotated Bibliography, International Reading Association, Newark, Delaware, 1980.

Mitchell, Graham, Kelly Carnes, and Cheryl Mendossa, "America's New Deficit: The Shortage of Information Technology Workers", Office of Technological Policy, U.S. Department of Congress, September 1998, http:// www.it.gov.doc [Accessed January 29, 1999].

Moore, Gary C. and Izak Benbasat, "Development of an Instrument to Measure Perceptions of Adopting an Information Technology Innovation", Information Systems Research, Vol. 2, No.3, 1991, p. 192-222.

Mylopolous, John, "Conceptual Modeling and Telos", in *Conceptual Modeling, Databases, and CASE: An Integrated View of Information Systems Development*, eds. Pericles Loucopoulos, Roberto Zicari, John Wiley & Sons, 1992, Chapter 2, pp. 49-68.

Norusus, Marja, *SPSS Advanced Statistics 6.1*, SPSS Inc. Chicago Illinois, 1994.

Nosek, John and Judith Ahrens, "An Experiment to Test User Validation of Requirements: Data Flow Diagrams vs. Task Oriented Menus", International Journal of Man-Machine Studies, 25, 1986, pp 675-84.

Oei, J.L.H., L.J. van Hemmen, E.D. Falkenberg, and S Brinkkemper, "The Meta Model Hierarchy: A Framework for Information Systems Concepts and Techniques", Technical Report, Department of Informatics, Faculty of Mathematics and Informatics, Katholieke Universiteir, Nijmegen, No. 92-17, July 1992, p. 1-30

Parsons. Jeff, and Yair Wand, "Using Objects for System Analysis", Communication of the ACM, Vol. 40, No. 12, December 1997, p. 104-110.

Ramsey, Rudy, Michael Atwood, and James Van Doren, "Flowcharts Versus Program Design Languages: An Experimental Comparison", Communications of the ACM, Vol. 26, No. 6, December 1993, p. 445-449.

Ramstad, Evan, “Software Spiral Means Breakthroughs, Glitches”, *Austin American-Statesman*, April 10, 1995, pp. C1.

“Rational”, Rational Software: Unified Modeling Language, http://www.rational.com, [Accessed May 25, 1999].

Robey, Daniel and Lynne Markus, “Rituals in Information System Design”, *MIS Quarterly*, Vol. 8, Number 1,, March 1984, pp. 5-15.

Rolland, C. and C. Cauvet, “Trends and Perspectives in Conceptual Modelling”, in *Conceptual Modeling, Databases, and CASE: An Integrated View of Information Systems Development*, eds. Pericles Loucopoulos, Roberto Zicari, John Wiley & Sons, 1992, Chapter 1, pp. 27-48.

Rossi, M and S. Brinkkemper, “Complexity Metrics for System Development Methods and Techniques”, Information Systems, 20, 1996 p. 53-63.

Siau, Keng Leng, *Empirical Studies in Information Modeling*, PhD. Thesis, University of British Columbia, September, 1996.

Siau, Keng and Matti Rossi, “Evaluation of Information Modeling Methods - A Review”, Proceeding 31st Annual Hawaii International Conference on System Science, ed. Daniel Dolk, Vol. V, p 314-322., 1998.

Siddiqi, Jawed, and M. Chandra Shekaran, “Requirements Engineering: The Emerging Wisdom, !EEE Software, March 1996, p.15-19

Sinha, Atish P. and Iris Vessey, “End-User Data Modeling: An Ontological and Empirical Approach of Relational and Object Oriented Models”, Working Paper, July 24, 1995, pp. 1-34.

SPSS, Statistical Software for Social Sciences, SPSS Inc. 444 N. Michigan Ave., Chicago, Illinois, Version 6.1and 8.0, 1998.

Stevens, James, “Applied Multivariate Statistics for the Social Sciences”, Second Edition, Lawrence, Erlbaum Associates, Hillsdale, 1992.

Stevens, Kevin T., Kathleen Stevens, and William Stevens, “Measuring the Readability of Business Writing: The Cloze Procedure Versus Readability Formulas”, The Journal of Business Communication, Vol. 29, No.4, p. 367-382.

Stewart, Thomas “Welcome to the Revolution”, *Fortune*, December 13, 1993, pp. 66-81.

Storey, V.C., “Relational Database Design Based on the Entity Relationship Model”, Data and Knowledge Engineering 7 (1991), p. 47-83.

Straub, Detmer W., “Validating Instruments in MIS Research”, Management Information Systems Quarterly, June 1989, 9. 147-169.

“Survey Uncovers High Costs of Computer Downtime” *PR Newswire*, July 28, 1992, Financial Section, pp.2.

Taylor, W.L., "Cloze Procedure: A New Tool for Measuring Readability", Journalism Quarterly, 33,1953 pp. 415-33.

Teorey, T.J., D. Jang, and J.P. Fry, "A Logical Design Methodology for Relational Databases Using the Extended Entity-Relationship Model," Computing Survey 18 (2), June, 197-221.
Todd, Peter and Izak Benbasat, "Process Tracing Methods in Decision Support Systems Research: Exploring the Black Box", Management Information Systems Quarterly, December 1987, pp. 493-512.

Todd, Peter and Izak Benbasat, "Process Tracing Methods in Decision Support Systems Research: Exploring the Black Box", Management Information Systems Quarterly, December 1987, pp. 493-512.

Vessey, Iris, Sirka Jarvenpaa, and Noam Tractinsky "Evaluation of Vendor Products: CASE Tools as Methodology Companions", Communications of the ACM, Vol. 35, No. 4, 1992, pp. 90-105.

Vessey, Iris and Sue Conger, "Requirements Specification: Learning Object, Process, and Data Methodologies", Communications of the ACM, May 1994, Vol. 37, No. 5, pp. 102-113.

Wand, Yair, " A Proposal for a Formal Model of Objects", Chapter 21, p. 537-539, in *Object Oriented Concepts, Databases, and Applications*, Kim, W. and F.H. Lanchovoy, eds., ACM Press, Addison Wesley Publishing Co., Reading, Massachusetts, 1989

Wand Yair, Andrew Gemino, and Carson Woo, "An Empirical Comparison of Object Oriented and Data Flow Modeling", Proceedings CAiSE "97, IFIP Workshop 8.1, Barcelona, Spain, June 16-20, 1997.

Wand, Yair, D.E. Monarchi, J. Parsons, and Carson Woo, "Theoretical Foundations for Conceptual Modeling in Information Systems Development", Decision Support Systems, 15, 1995 , p. 285-304.

Wand, Yair, Veda Storey, and Ron Weber, "An Ontological Analysis of the Relationship Construct in Conceptual Modeling", Working Paper 97-MIS-005, Faculty of Commerce and Business Administration, University of British Columbia, February1999 (revised).

Wand, Yair and Ron Weber, "On the Ontological Expressiveness of Information Systems Analysis and Design Grammars", Journal of Information Systems, 1993,, vol. 3, pp. 217-237.

Wand, Yair and Ron Weber, "On the Deep Structure of Information Systems", Journal of Information Systems, vol. 5, 1995, p. 203-223.

Wand, Yair and Carson Woo, "Object Oriented Analysis - Is it Really That Simple?", Proceedings of the Workshop on Information Technologies and Systems, Orlando Florida, December 4-5, 1993, p. 186-195.

Yadav, Surya , Ralph Bravococ, Akem Chatfield, and T. Rajkumar, "Comparison of Analysis Techniques for Information Requirements Determination", Communications of the ACM, Sept. 1988, Vol. 31, Number 9, pp. 1090-97.

Appendix A: Text Description and Diagrams for Cases

Organizing an IFIP Conference

An IFIP Working Conference is an international conference centered on a topic of specific interest to one or more IFIP Working Groups. Participation in the conference is by invitation only. Two objectives of the conference organizers are to ensure that members of the involved Working Group(s) and Technical Committee(s) are invited and that attendance is sufficient for financial break-even without exceeding the capacity of the facilities available.

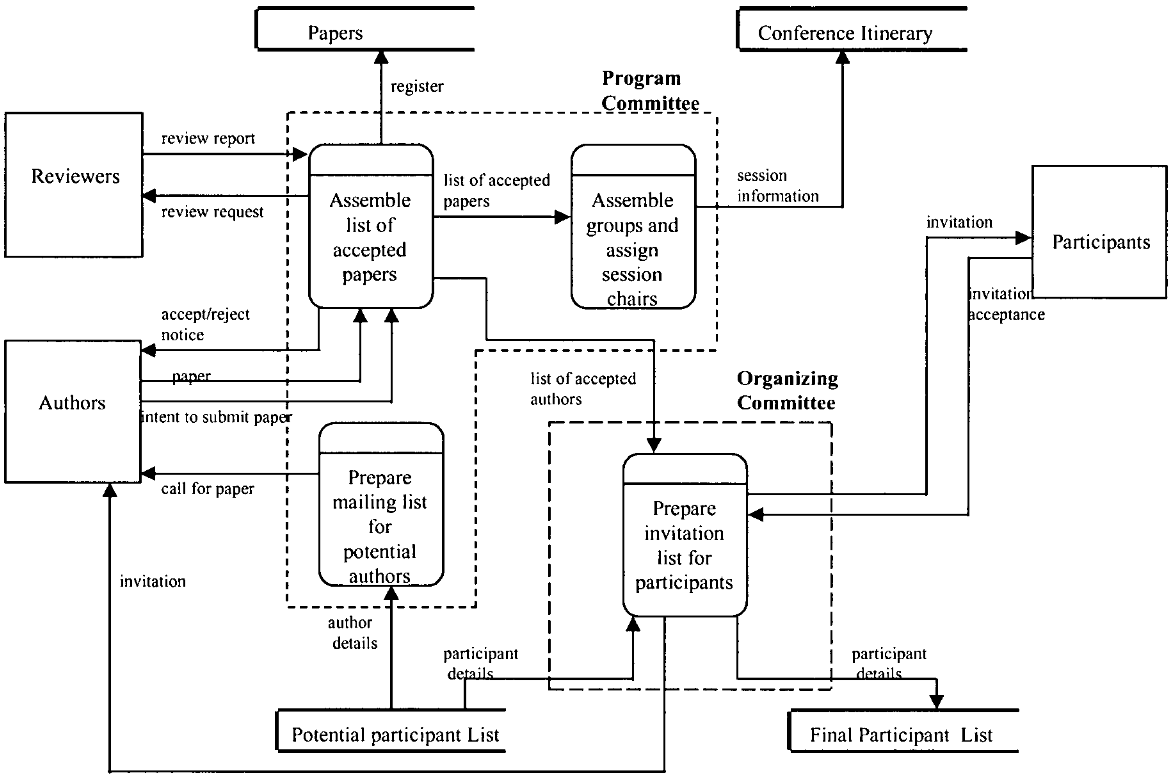
Two committees are involved in organizing an IFIP Working Conference: a Program Committee and an Organizing Committee. The Program Committee deals with the technical content of the conference and the Organizing Committee handles financial and local arrangements, along with invitations and publicity. These committees work together closely and have a need for common information.

When an IFIP conference is to be held, the Program Committee is responsible for several activities. These include preparing a mailing list of potential authors and sending a call for papers to each of these individuals. If the potential authors reply with a letter of intent indicating they will submit a paper to the conference, the Program Committee registers the participant's intent to participate. When the paper is received, the paper is also registered.

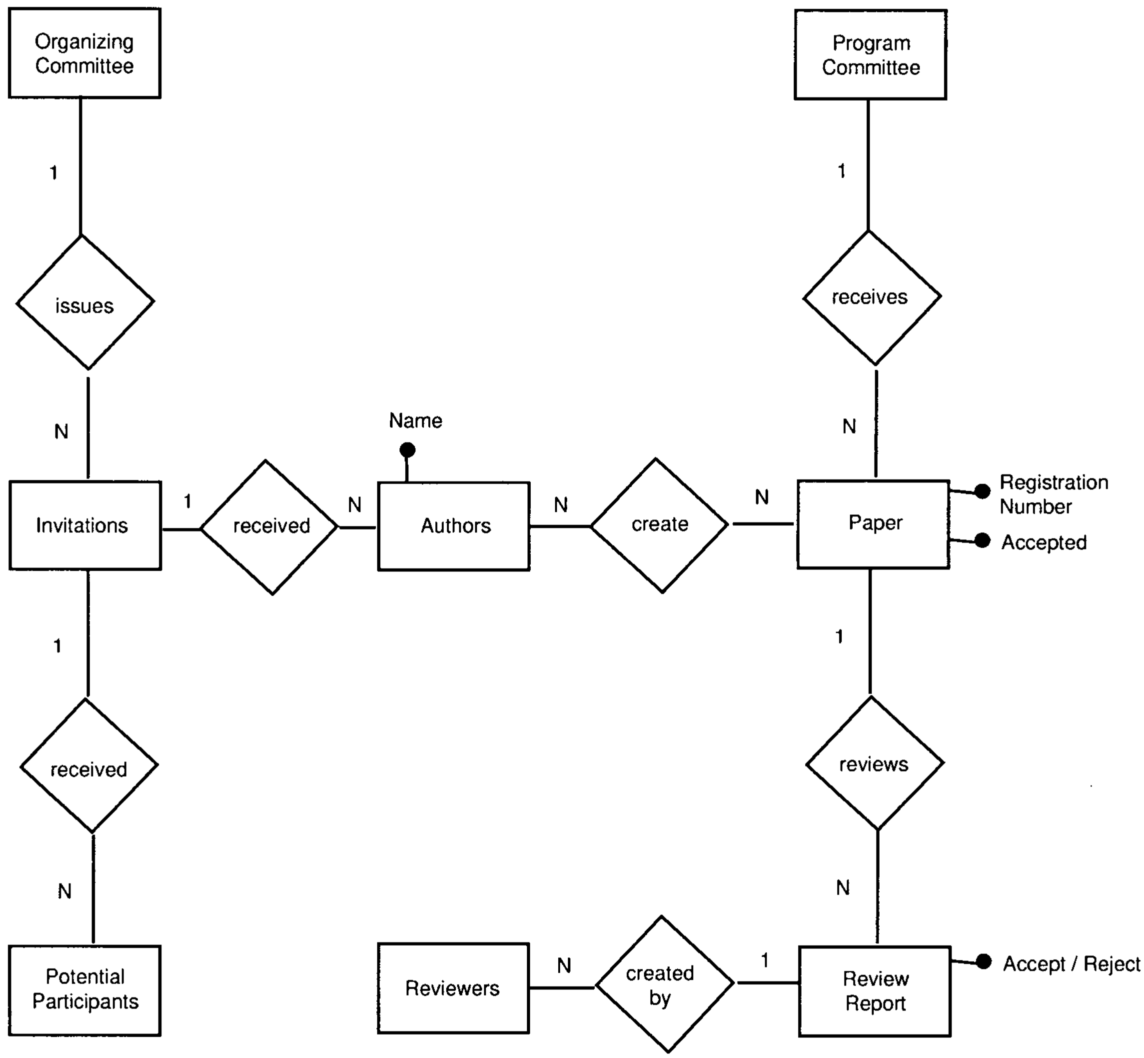
After receiving the papers, the Program Committee assigns a set of reviewers for each paper and then sends the papers to the respective reviewers. Reviewers create a review report and send their reports to the committee. The committee then groups accepted papers into sessions and assigns session chairs. The list of accepted authors is forwarded to the organizing committee. This session information makes up the itinerary for the conference.

The Organizing Committee begins by preparing a list of people to invite to the conference. The committee issues priority invitations to National Representatives and to members of related Working Groups. The organizing committee also ensures that each of the contributing authors receives an invitation. Individuals who receive an invitation and who intend to come to the conference must indicate their intent to participate by sending an acceptance of invitation to the Organizing Committee. The final list of participants is then generated by the Committee, who makes an effort to avoid sending duplicate invitations to any individual.

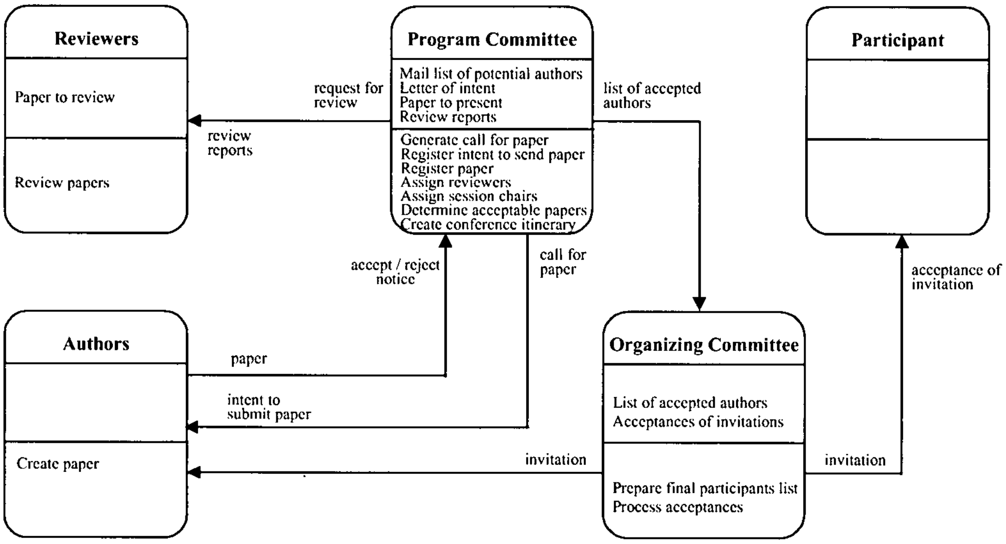
Organizing an IFIP Conference



Organizing an IFIP Conference



Organizing an IFIP Conference



Organizing an Entertainment Event

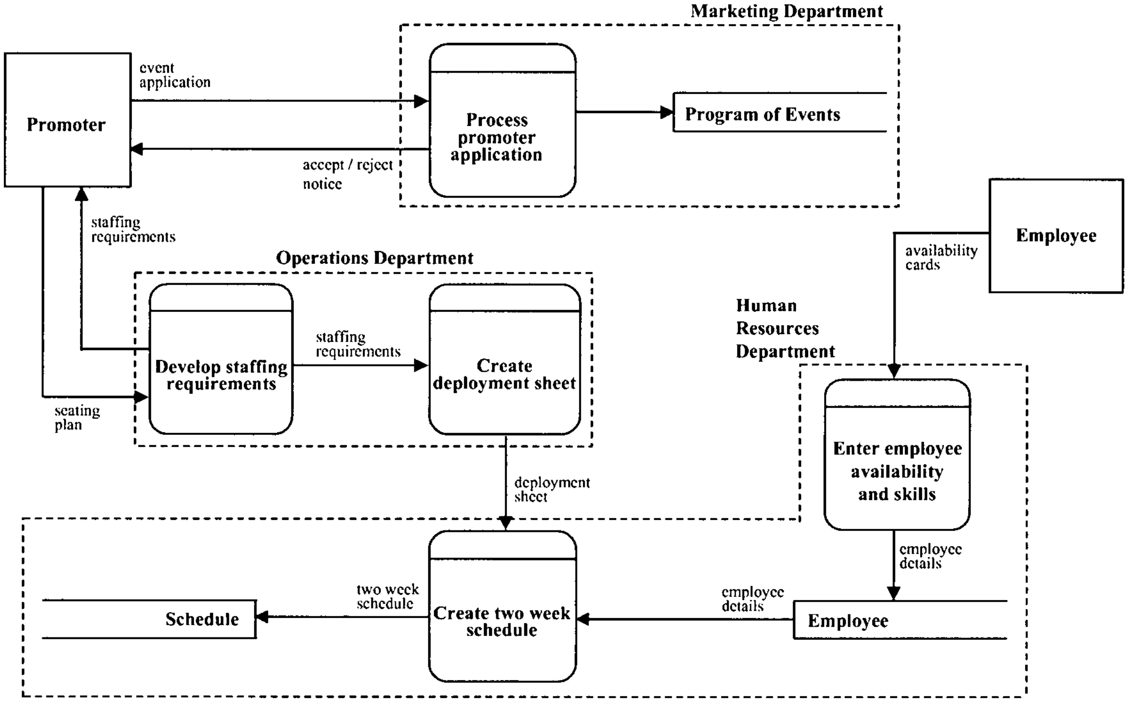
A company that owns a large sports and entertainment complex holds a variety of events in their building including professional sports, concerts, and conventions. The program of events is planned several months in advance. While no more than one event can take place at any given time, it is possible for more than one event to occur on the same day.

To book an event, a promoter contacts the marketing manager for the complex and explains the type of event being promoted, the expected attendance, and a requested date and time for the event. After reviewing the promoter's application, the marketing manager either accepts the application and places the event in the program of events, or rejects the application. Once an event application is accepted, the promoter is responsible for sending a seating plan for the event to the operations manager of the complex.

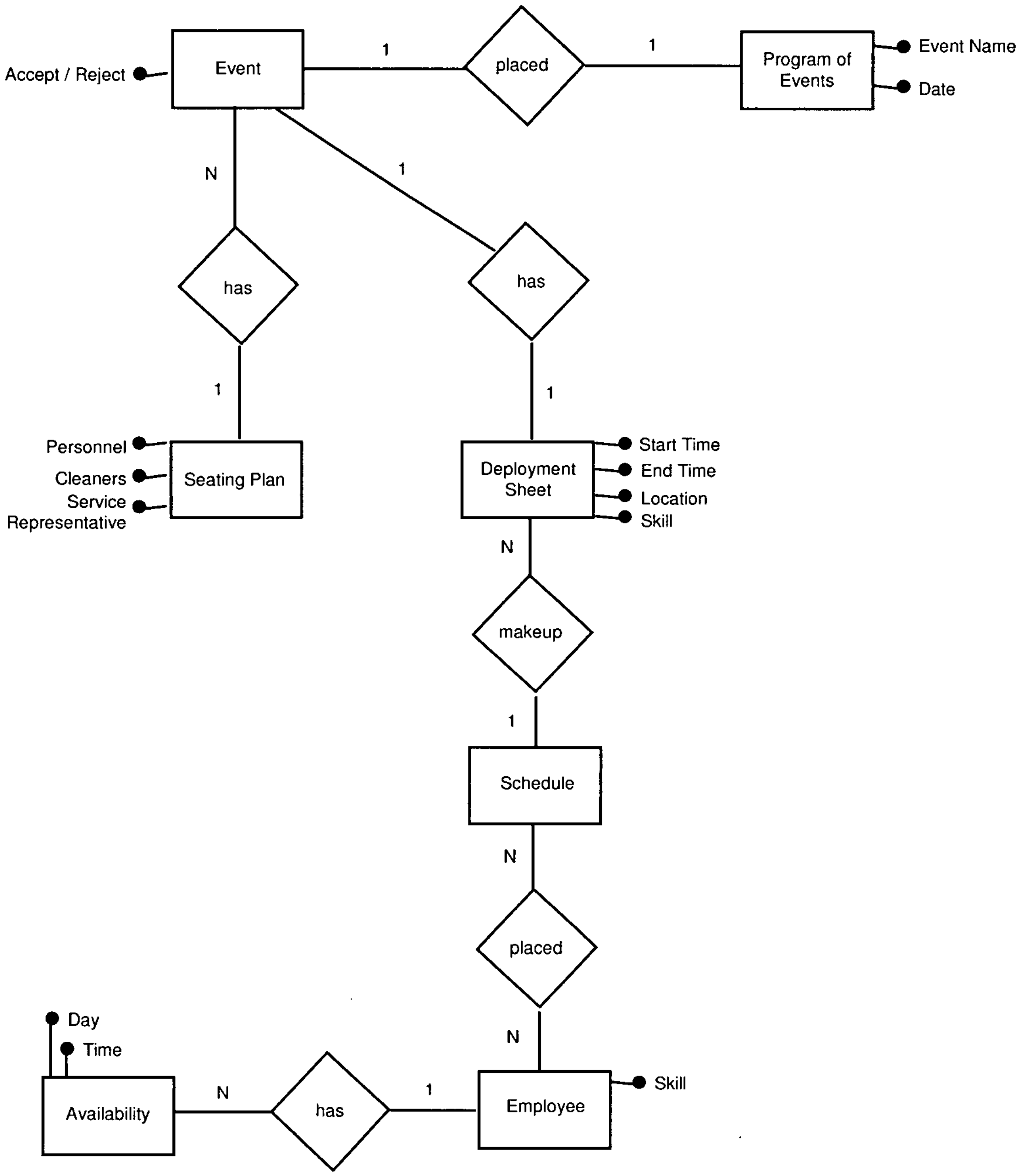
The operations manager reviews all seating plans and decides on the appropriate number of security personnel, cleaners, and customer service representatives for the event. These estimates for staffing requirements are reviewed with the promoter before they are finalized. Once the staffing requirements are finalized, a deployment sheet is created by the operations manager. The deployment sheets lists the start time, end time, location, and skill requirements for every shift that will be required for the event. The completed deployment sheets are then passed to the human resource manager who is responsible for scheduling staff to work the event.

The human resource department creates a schedule two weeks in advance for all of the events to be held in the upcoming two week period. The deployment sheets from all of the events in the two week period are used to assign shifts to the employees. The scheduler works through an entire deployment sheet one shift at a time. Employees are required to fill out an employee availability card indicating their availability and skills. No employee can work over 30 hours a week. In order for the scheduler to assign a shift to an employee, the employee must have the right skills and be available for the event. When the scheduler is finished this process, the final schedule is posted for the employees.

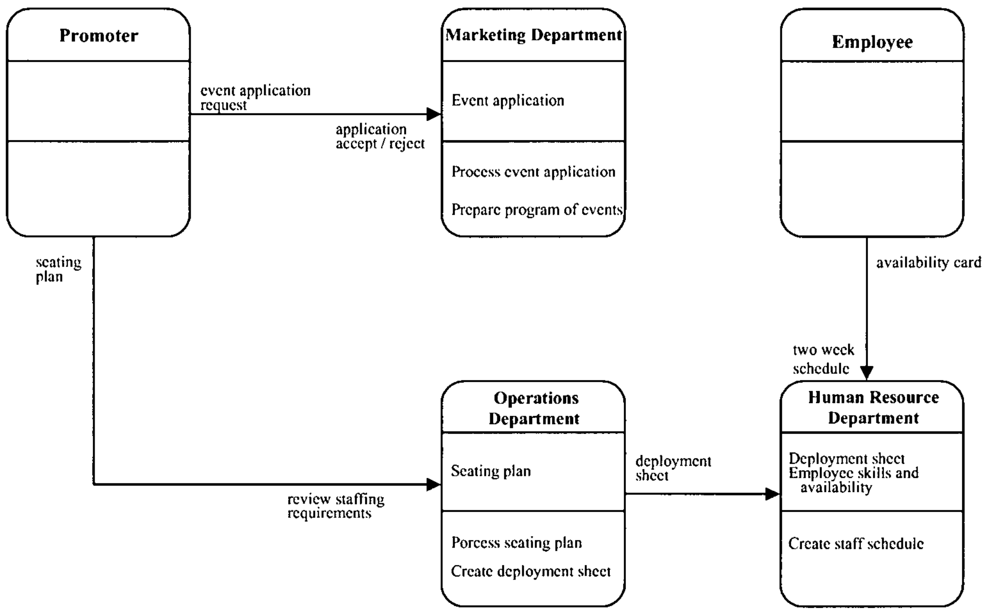
Organizing an Entertainment Event



Organizing an Entertainment Event



Organizing an Entertainment Event



Voyager Bus Inc.

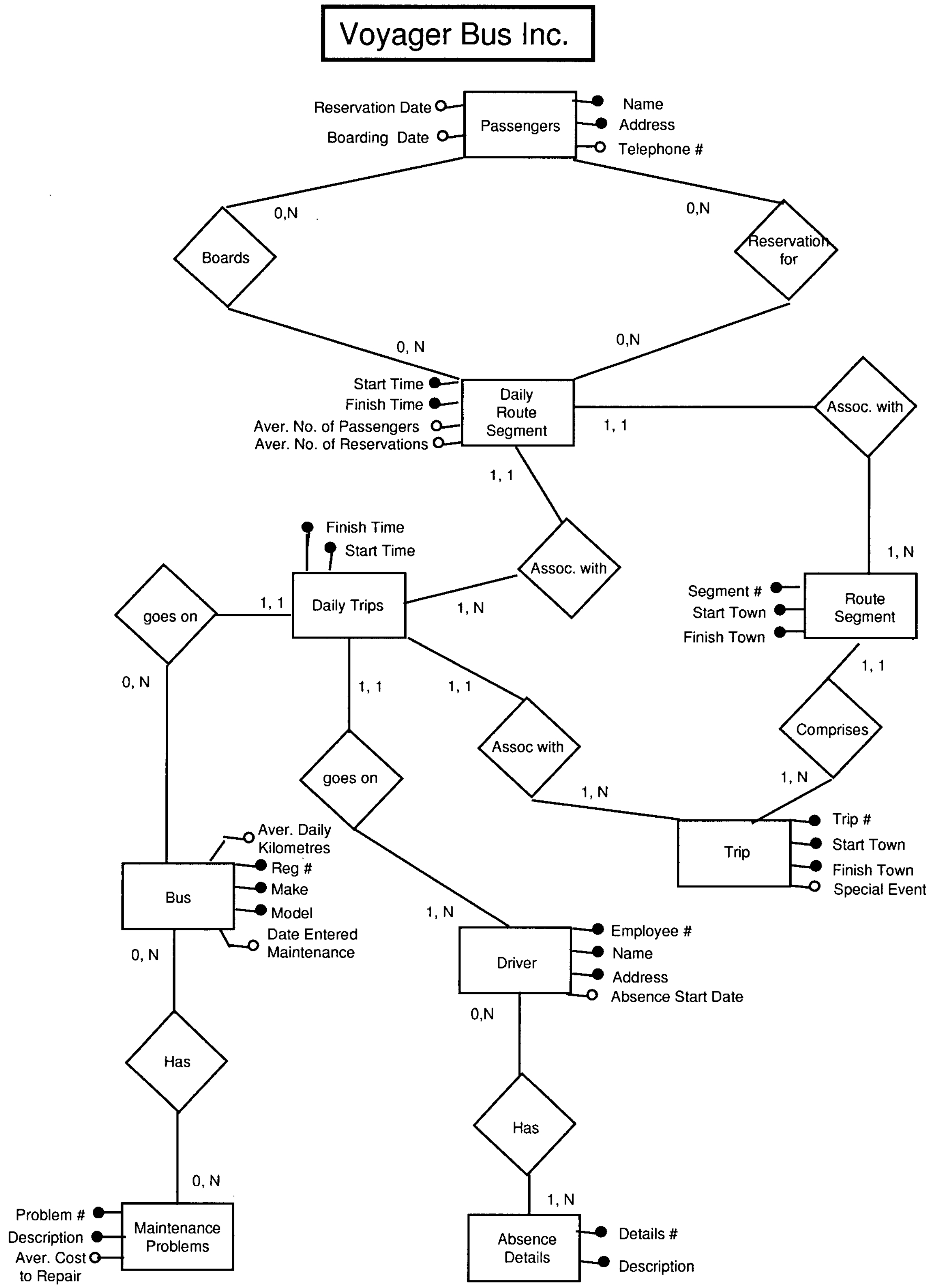
Voyager Bus Inc. (Voyager) is a company specializing in bus trips to places of interest or special events. The objectives of Voyager are to provide high quality and safe traveling experiences for tourists.

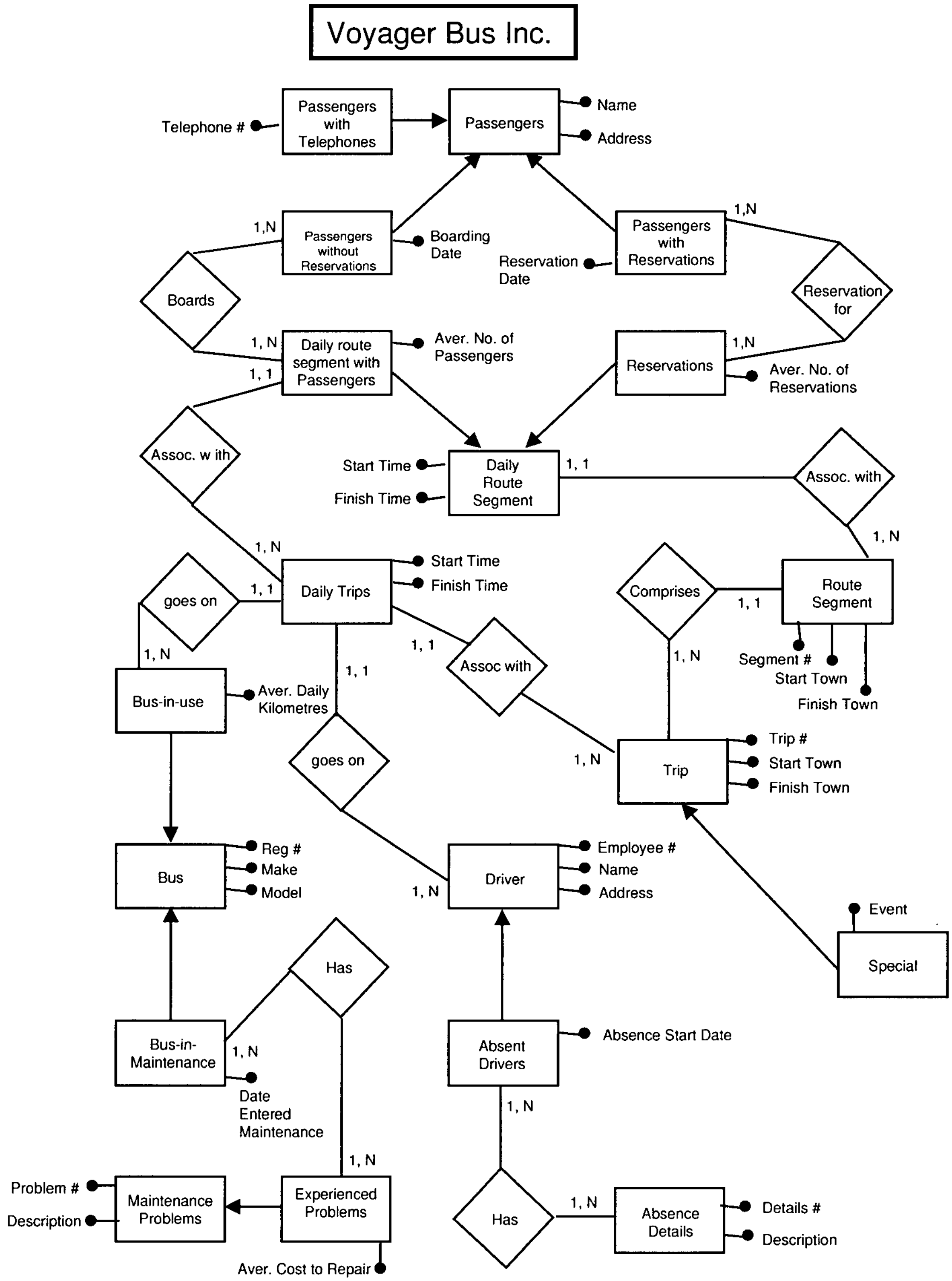
There are two ways for people to travel with Voyager. Passengers can either make a reservation on a trip, or passengers can show up at the boarding gate without a reservation and purchase a ticket for an unreserved seat. Passengers with a reservation are assigned a reservation date, whereas, passengers without reservations are assigned a boarding date. The name and addresses of all passengers are collected. Telephone numbers are collected where possible.

All bus trips are organized into daily route segments. All daily route segments have both a start time and an end time. Each daily route segment Voyager organizes is classified as a route segment with a segment number, start town, and finish town. Voyager offers a range of trips, and each trip is made up of one or more route segments. For every trip there is a trip number, start town, and finish town. If the trip is organized around a special event, the event name is also associated with the trip.

Each daily route segment that Voyager offers is part of a daily trip. A daily trip is undertaken by one or more bus drivers. The name, address, and employee number of all drivers is collected. Voyager also records information about absent drivers. When a driver is absent, Voyager records the absence start date and the details about the absence. The absent driver provides one or more reasons for being absent and each reason is assigned a detail number and a short description.

Voyager also collects information about the buses used for daily trips. Buses have a make, model, and registration number. For buses in use, the average daily kilometers is collected. If a bus requires maintenance, Voyager notes the date on which the bus entered maintenance and records the one or more problems with the bus. Voyager assigns a problem number and a short description for every maintenance problem. Finally, the average cost to repair all problems with a bus in maintenance is also recorded.


Voyager Bus Inc.
Passengers
Reservation Date
Boarding Date
Name
Address
Telephone #
Boards
Reservation for
0,N
0,N
0, N
0,N
Daily Route Segment
Start Time
Finish Time
Aver. No. of Passengers
Aver. No. of Reservations
1, 1
Assoc. with
1, 1
Assoc. with
1, N
Daily Trips
Finish Time
Start Time
1, N
goes on
1, 1
0, N
Route Segment
Segment #
Start Town
Finish Town
1, 1
Comprises
1, N
1, 1
1, 1
Assoc with
1, N
goes on
Trip
Trip #
Start Town
Finish Town
Special Event
Bus
Aver. Daily Kilometres
Reg #
Make
Model
Date Entered Maintenance
1, N
Driver
Employee #
Name
Address
Absence Start Date
0, N
0,N
Has
Has
0, N
1, N
Maintenance Problems
Problem #
Description
Aver. Cost to Repair
Absence Details
Details #
Description


Voyager Bus Inc.
Telephone #
Passengers with Telephones
Passengers
Name
Address
1,N
Passengers without Reservations
Boarding Date
Passengers with Reservations
Reservation Date
1,N
Boards
Reservation for
1, N
Daily route segment with Passengers
Aver. No. of Passengers
1, 1
Reservations
1,N
Aver. No. of Reservations
Assoc. w ith
Start Time
Finish Time
Daily Route Segment
1, 1
Assoc. with
1, N
1, N
Daily Trips
Start Time
Finish Time
goes on
1, 1
Comprises
Route Segment
1, 1
1, N
Segment #
Start Town
Finish Town
1, N
1, 1
Assoc with
1, 1
Bus-in-use
Aver. Daily Kilometres
goes on
1, N
Trip
Trip #
Start Town
Finish Town
Bus
Reg #
Make
Model
1, N
Driver
Employee #
Name
Address
Event
Special
Has
Bus-in-Maintenance
1, N
Date Entered Maintenance
Absent Drivers
Absence Start Date
1, N
1, N
Problem #
Description
Maintenance Problems
Experienced Problems
Aver. Cost to Repair
Has
1, N
Absence Details
Details #
Description

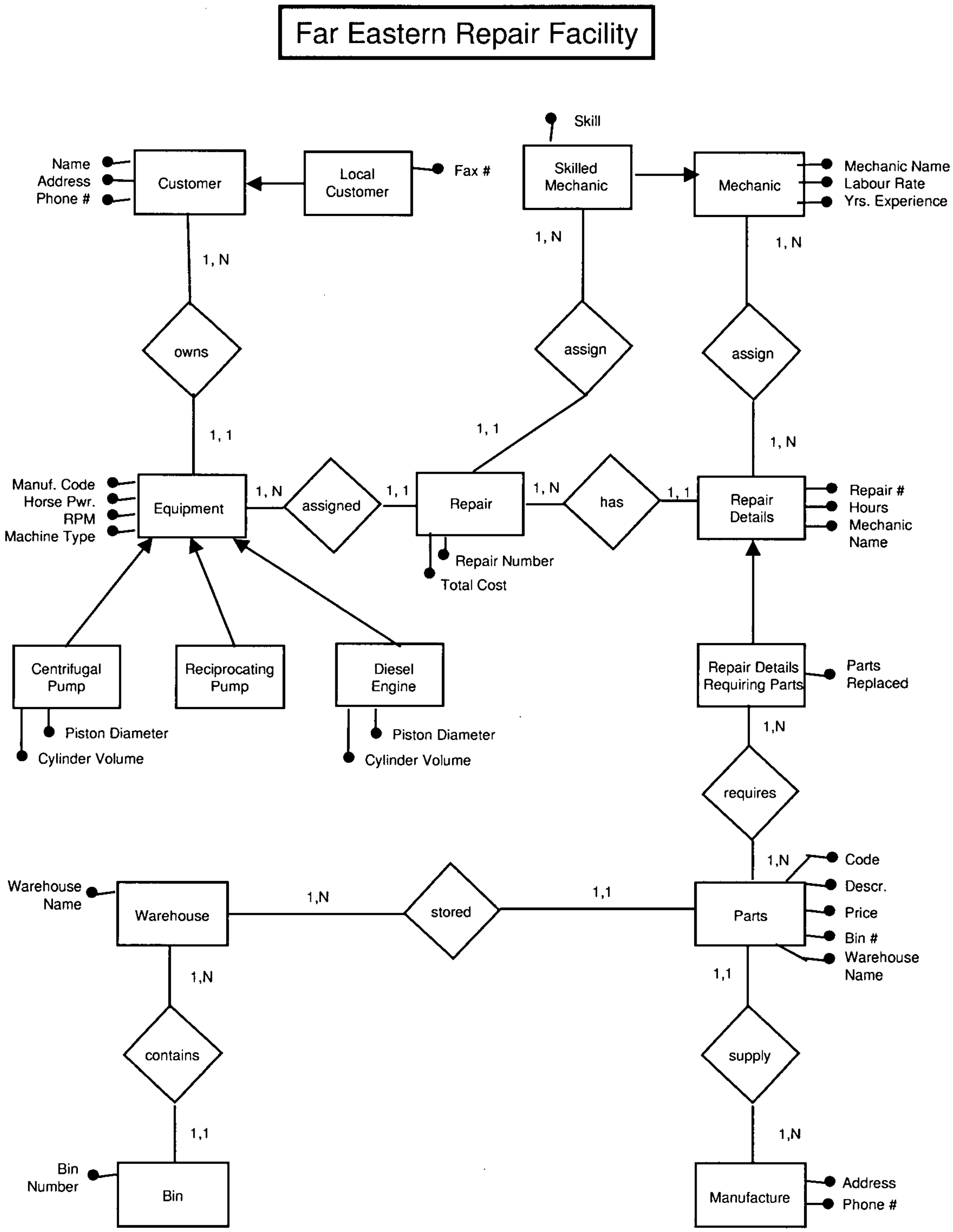
Far Eastern Repair Facility

The Far-Eastern Repair Facility carries out repairs of manufactured equipment for clients. The objectives of the Far Eastern Repair Facility (Far Eastern) are to run an efficient repair facility that provides high quality repair service in the shortest possible time. The repair facility has provided Far Eastern with the capability to repair three general types of manufactured equipment: centrifugal pumps, reciprocating pumps, and diesel engines.

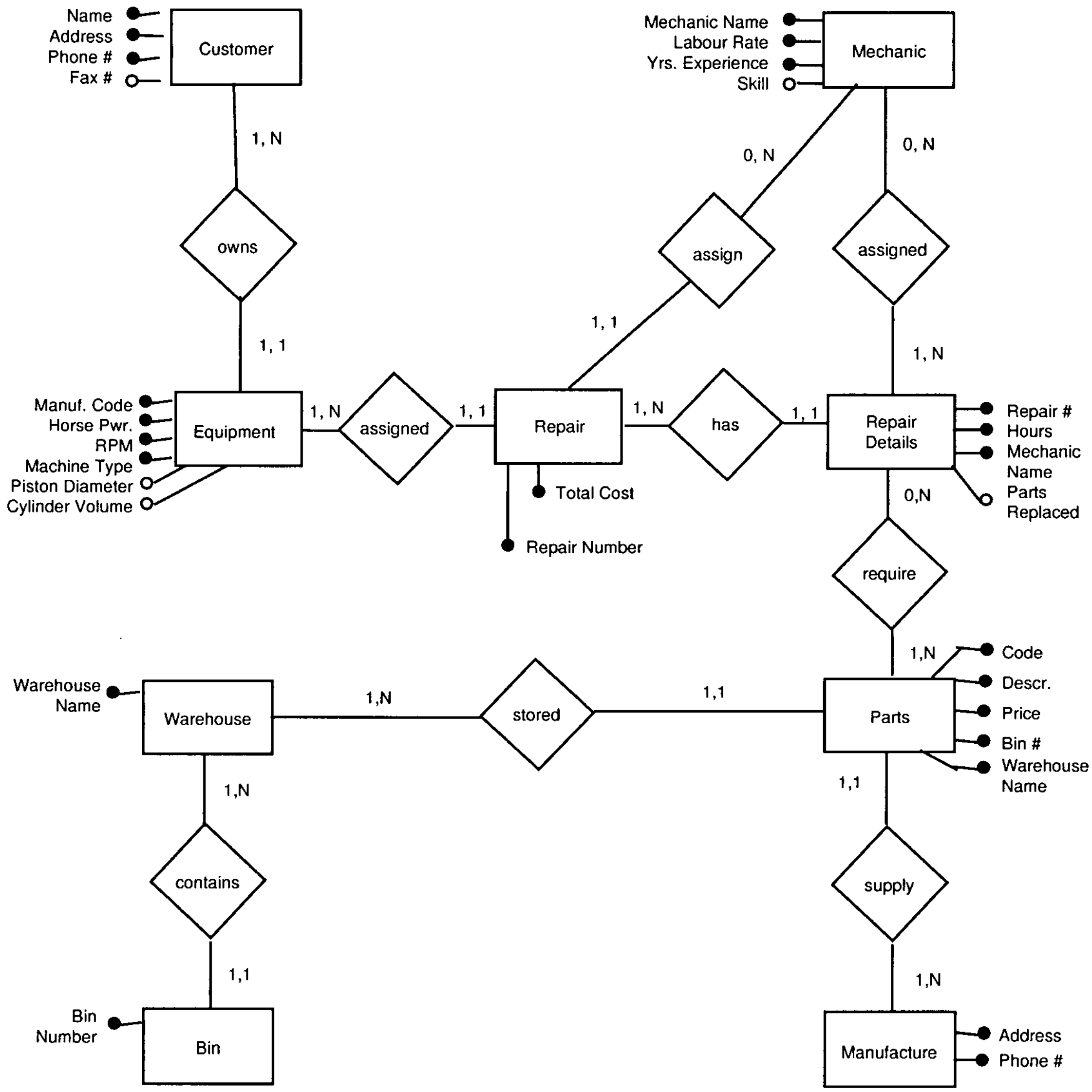
When Far Eastern receives a piece of equipment in need of repair, the company assigns a repair number and records the original equipment manufacturer's code along with the horsepower and speed (RPM) at which the machine will run. For centrifugal pumps and diesel engines, Far Eastern also records both the piston diameter and cylinder volume associated with each piece of equipment. Every piece of equipment brought to Far-Eastern Repair is owned by a customer. Far Eastern maintains the current address, and phone number of the customers for billing purposes. Fax numbers are collected only for local customers.

Far Eastern maintains an inventory of spare parts. The repair facility collects the address and phone number of each part manufacturer so that they can order spare parts for machines they repair. The parts inventory is warehoused in three different buildings labeled warehouse 1, 2, and 3. Parts are stored in numbered bins within each warehouse. Each part is identified by a part code along with the description, list price, bin number, and weight for each part.

There are several mechanics at the facility. Since each mechanic differs in skill and experience, each mechanic has a different labor rate. If a mechanic has a special skill, that skill is recorded by Far Eastern. Years of experience is also recorded. When equipment arrives for repair, one of the skilled mechanics are assigned to the repair task. The skilled mechanic then can assign other mechanics to details associated with the repair. When the repair is completed, each mechanic working on the repair task enters the number of hours they spent on repair, parts replaced (if any), and a description of the repair. The total cost associated with the repair task is then calculated and recorded.


Far Eastern Repair Facility
Skill
Name
Address
Phone #
Customer
Local Customer
Fax #
Skilled Mechanic
Mechanic
Mechanic Name
Labour Rate
Yrs. Experience
1, N
1, N
1, N
owns
assign
assign
1, 1
1, 1
1, N
Manuf. Code
Horse Pwr.
RPM
Machine Type
Equipment
1, N
assigned
1, 1
Repair
1, N
has
1, 1
Repair Details
Repair #
Hours
Mechanic Name
Repair Number
Total Cost
Centrifugal Pump
Reciprocating Pump
Diesel Engine
Repair Details Requiring Parts
Parts Replaced
Piston Diameter
Cylinder Volume
Piston Diameter
Cylinder Volume
1,N
requires
1,N
Code
Descr.
Price
Bin #
Warehouse Name
Warehouse Name
Warehouse
1,N
stored
1,1
Parts
1,N
1,1
contains
supply
1,1
1,N
Bin Number
Bin
Manufacture
Address
Phone #

Far Eastern Repair Facility



Appendix B: Pretest Questions

Pretest: Knowledge of Analysis Methods - All Cases

Prior Use of Analysis Methods

Have you ever used data Flow Diagrams to Model a Business Organization?	Y / N
Have you ever used Object Oriented Diagrams to Model a Business Organization?	Y / N
Have you ever used Entity Relationship Diagrams to Model a Business Organization?	Y / N

Familiarity with Analysis Methods

For how many months have you been familiar with Data flow Diagrams?	0 - 36
For how many months have you been familiar with Object Oriented Diagrams?	0 - 36
For how many months have you been familiar with Entity Relationship Diagrams?	0 - 36

Competence with Analysis Methods

Data flow Diagrams	0 - 7
Object Oriented Diagrams	0 - 7
Entity Relationship Diagrams	0 - 7

Confidence in Analysis Methods

Data flow Diagrams	0 - 7
Object Oriented Diagrams	0 - 7
Entity Relationship Diagrams	0 - 7

Pretest: Knowledge of Domain - IFIP Conference Case

Please indicate your level of knowledge of the following businesses:

Organizing an academic conference	0 - 7

Please indicate which of the activities listed below you have done:

Attended an academic conference	Y / N
Helped to organize a conference	Y / N
Reviewed a paper for a conference	Y / N
Submitted a paper for a conference	Y / N
Created a mailing list for a large group	Y / N

Pretest: Knowledge of Domain - Entertainment Event Case

Please indicate your level of knowledge of the following businesses:

Organizing an entertainment event	0 - 7

Please indicate which of the activities listed below you have done:

Helped to organize an entertainment event	Y / N
Worked at a large entertainment event	Y / N
Promoted an entertainment event	Y / N
Delegated work for a group of employees	Y / N
Arranged security for an entertainment event	Y / N

Pretest: Knowledge of Domain - Voyager Bus Inc.

Please indicate your level of knowledge of the following businesses:

Organizing a bus tour company	0 - 7

Please indicate which of the activities listed below you have done:

Taken a Bus tour	Y / N
Worked as a bus driver	Y / N
Made a reservation for a bus trip	Y / N
Traveled by bus to a special event	Y / N
Organized a set of short bus trips	Y / N

Pretest: Knowledge of Domain - Far-Eastern Repair Facility

Please indicate your level of knowledge of the following businesses:

Organizing a machine repair facility	0 - 7

Please indicate which of the activities listed below you have done:

Worked as a mechanic	Y / N
Worked in a warehouse	Y / N
Replaced a part on an engine	Y / N
Had your engine overhauled	Y / N
Helped to organize a repair shop	Y / N

Appendix C: Comprehension Questions

Organizing an IFIP Conference

Question		Answer
1.	Is the organizing committee responsible for preparing the final list of participants?	Y
2.	Does an author know if their paper has been accepted by the program committee?	Y
3.	Is the organizing committee responsible for developing the conference itinerary?	N
4.	Does a participant have to be an author to be placed on the final list of participants, or are there other ways to participate?	N
5.	Is the program committee responsible for maintaining a list of accepted authors?	Y
6.	Does the program committee offer invitations to participants?	N
7.	Do the Program Committee and Organizing Committee have to interact with each other?	Y
8.	Does the program committee keep track of all the papers that are submitted to the conference?	Y
9.	Are all of the papers that are submitted to the program committee accepted by the program committee?	N
10.	Can an author also be a part of the organizing committee?	U
11.	Does the program committee have access to the final list of participants?	N
12.	Is a paper the first thing that an author sends to the program committee?	N

Organizing an Entertainment Event

Question		Answer
1.	Is the operations department responsible for deciding when events are to be held at the complex?	N
2.	Does a promoter know that their event application has been accepted and placed into the program of events?	Y
3.	Is the human resource department responsible for creating the deployment sheets?	N
4.	Can an employee be scheduled on days they are not available?	N
5.	Is the promoter the person who determines the number of employees who should work an event?	N
6.	Is the operations department responsible for creating the employee schedule?	N
7.	Do the marketing department and the operations department share similar information?	Y
8.	Is the marketing department responsible for keeping track of all of the events that will be held in the upcoming months?	Y
9.	Are all of the seating plans provided by the promoter accepted by the operations department?	U
10.	Do all employees work every event?	N
11.	Is the operations department responsible for creating a seating plan for every event?	N
12.	Is the seating plan directly used in the development of the employee schedule?	U

Voyager Bus Inc.

Question		Answer
1.	Can a trip be made up of more than one route segment?	Y
2.	Does a person have to make a reservation to go on a trip?	N
3.	Can a daily trip be assigned to more than one bus?	Y
4.	Does Voyager Bus Inc. collect the same set of information for all of the passengers?	N
5.	Can the same daily route segment be associated with two different trip numbers?	Y
6.	Can Voyager Bus Inc. record a maintenance problem that has not yet been experienced by any of their buses?	U
7.	Is the daily route segment modeled as an entity?	Y
8.	Can a bus driver be assigned to more than one daily trip?	Y
9.	Are all buses that are available for use assigned to a daily route segment?	U
10.	Is model an attribute of bus?	Y
11.	Is the average cost of repair recorded for all buses in maintenance?	Y
12.	Can the end town assigned to a route segment be different from the end town associated with a trip that uses the route segment?	Y

Far-Eastern Repair Facility

Question		Answer
1.	Do all repairs require parts?	N
2.	Can a repair be worked on by more than one mechanic?	Y
3.	Are all repairs assigned to at least one mechanic	Y
4.	Are there parts stored in the warehouse that are not used for repairs?	U
5.	Does Far Eastern collect different information for different machine types?	Y
6.	Does Far Eastern differentiate their local customers in any way?	Y
7.	Can a mechanic who does not have a special skill be assigned to more than one repair?	Y
8.	Do all the mechanics related to the same repair, pool their hours to create a single entry for hours worked?	N
9.	Can a piece of equipment undergo more than one repair?	Y
10.	Can more than one part be listed in a single repair detail?	Y
11.	Is the cylinder volume recorded for all pumps that are repaired?	N
12.	Can a part be supplied by more than one manufacturer?	Y

Appendix D: Problem Solving Questions and Answers

Problem Solving Questions and Answers: IFIP Conference Case

Question 1:
Is it possible for the same individual to receive two invitations to an IFIP conference?

If you answer yes, indicate how this duplication might occur. (Provide as many suggestions as you can think of.) If you answered no, indicate why it cannot happen. (Provide as many suggestions as you can think of.)

Answers

1. Yes. An individual could be both in the potential participation list and the list of accepted authors.
2. Yes. Duplication in the potential participation list.
3. Yes. Duplication in potential authors participation list
4. Yes. An author may have more than one paper.
5. No. If invitation is sent out at end of process (after list of accepted authors), then no duplication.
6. Individual was a special representative and an author.
7. Individual was a national representative and part of a related working group.

Question 2:
Suppose that a person was invited to submit a paper but has not yet received an invitation. What could have happened? (Provide as many possible suggestions as you can think of).

Answers

1. Person's paper was rejected.
2. Invitation was supposed to be sent but there was a delay in processing.
3. Person did not submit an "intent to submit paper" (no registration occurred).
4. Author has moved and conference mailing list has not been changed.
5. Reviewer did not return the paper.
6. The person never submitted a paper.
7. Program committee has not sent his name to the organizing committee.
8. The paper is still under review.
9. Paper was lost during review.
10. Person sent a note indicating they could not attend.
11. Submitted paper after deadline.
12. Two authors with the same name.

Question 3:
Neither the program committee nor the organizing committee is directly responsible for maintaining the mailing list of potential authors and participants. What organization problems do you think could arise from this situation? (Provide as many possible suggestions as you can think of).

Answers

1. Inconsistencies between two mailing lists (organizing and Program committee lists).
2. Invitations sent to incorrect addresses. This may lead to poor turnout.
3. Redundancy (duplication) across mailing lists. Likely to invite same people twice.
4. Miss potential participants by assuming they are present on the other mailing list.
5. Receive information twice.

Question 4:
Suppose that a very highly respected author indicated that she would like to present a paper in the conference but she has given the program committee only two weeks notice. What problems would be faced by the committees in accepting the author's wish? (Provide as many possible suggestions as you can think of).

Answers
1. Not enough time for review process.
2. Not enough time to generate invitation from organizing committee.
3. Requires change in the conference itinerary (which have been printed).
4. No time left in the conference itinerary.
5. Notices of change will have to be sent.
6. Sets a dangerous precedent. Other authors may want to do the same.
7. There is no process in place to support this (no exception handling).
8. Modify the list of participants.
9. No time to promote the author's presence at the conference.
10. Printed material must be changed.
11. Finding reviewers at short notice.

Question 5:
Suppose that instead of inviting specific authors to submit papers and inviting only selected participants to come to the conference, the two committees agreed to open participation to any individuals and all interested authors. What problems would arise if the current system was used to organize the conference? (Provide as many suggestions as you can think of).

Answers
1. Keep track of number of participants (by registering those interested).
2. Reviewers may be overwhelmed with papers. Screening papers may be necessary.
3. Would have to advertise for authors.
4. Would have to advertise for participants.
5. Would no longer need to send invitations or compile invitation list.
6. Larger event to organize.
7. Ticketing arrangement would be needed.
8. Two committees may not be necessary.
9. Keep track of accepted authors.
10. Budgeting for conference (how many?).
11. Loss of prestige.

Problem Solving Questions and Answers: Entertainment Event Case

Question 1:
Suppose that an employee was available to work on a Saturday but they were not scheduled to work. How could this have occurred? (Provide as many possible suggestions as you can think of).

Answers
1. The employee does not have the correct skills.
2. Staffing requirements did not need that many people (some left out).
3. Employee has worked too much this week.
4. There was no event on Saturday.
5. A mistake was made by human resources.
6. Employee data not up-to-date.
7. Employee did not turn in availability card.
8. Large number of staff available on Saturday.
9. Hours of availability restricted on Saturday.
10. Incorrect availability card.
11. Event cancellation or change.
12. Not listed on staff schedule.
13. Has worked up to his limit in hours.

Question 2:
Suppose that it is a very busy two weeks for the sport complex with events scheduled for every day of the week. What problems might be encountered due this busy schedule. (Provide as many possible suggestions as you can think of).

Answers
1. Scheduling conflicts.
2. Not enough employees for staffing.
3. Complex setup problems (ice to basketball).
4. Time needed to clean.
5. Might run out of supplies.
6. Crowding entertainers (not enough change rooms).
7. Time needed to create and adjust deployment sheets.
8. Impossible to extend events.
9. Hard to find employees at short notice in case of emergency.
10. Marketing manager may have trouble scheduling events.
11. Overtime for employees.
12. Audience may be too small or too big.
13. Weather causes delay.

Question 3:
Suppose that a promoter underestimated the number of people who wanted to attend a event. The promoter now wants to increase the number of seats available for an event. Given that there are some seats that could be opened, what problems would be created by opening more seats? (Provide as many possible suggestions as you can think of).

Answers
1. Extra tickets must be sold.

2. Staffing requirements may have to change.
3. Informing operations manager and providing adequate time for changes.
4. Promoter must provide new seating plan.
5. Deployment sheets have to change.
6. Staff schedule has to change (more employees required).
7. Cost increase due to overtime.
8. Employees added to schedule need notification.
9. Not enough employees.
10. Need to provide extra supplies.
11. Conflict with following event.
12. Confusion during seating.

Question 4:
Some events take a long time and a large number of people to setup and take down (this is referred to as conversion time). What organizational problems can be caused by events with a long conversion time? (Provide as many possible suggestions as you can think of)

Answers
1. More staffing required for big conversion events.
2. Loss in revenue due to conversion time.
3. Less time available to marketing for scheduling events.
4. Spend more money on overtime for employees.
5. Events may have to be delayed due to conversion time.
6. Stress on employees.
7. Shortage of staff.
8. Budgeting allocation is more complex.

Question 5:
Assume that the marketing department, operations department, and the human resources department act independently of each other. What problems might occur if the departments do not share their information quickly or work closely with the other departments? (Provide as many possible suggestions as you can think of).

Answers
1. Marketing schedules events that don't fit in complex.
2. Operations may under /overstaff an event.
3. HR department may not provide correct number of workers.
4. HR department may provide workers with the wrong skill.
5. HR department may hire the wrong skill type.
6. HR does not have deployment sheet to create staff schedule.
7. Departments may not be working with up-to-date information since they are unaware of updating process.
8. More expenses may be incurred by lack of communication.
9. Operations may start working on staffing requirements, even if event is not yet accepted.
10. Employees may be scheduled for an event that has been canceled.
11. An event that cannot be staffed may not be canceled in time.
12. Staff schedule may not reflect last minute changes.
13. Marketing may not give information regarding events to operations, so no deployment sheet is created.
14. Public relations may be affected due to inadequate response time between departments.

Problem Solving Questions and Answers: Voyager Bus Inc.

Question 1:
An employee at Voyager Bus Inc. has come up with an idea for a new trip, and has assigned the trip with a trip number, start town, and end town. The employee showed the newly planned trip to his manager and she said that the trip, although exciting, was not possible. What reasons can you provide for the trip not being possible? (Provide as many solutions as you can think of.)

Answers
1. Not able to fit trip to daily route segments.
2. Not enough buses.
3. Not enough drivers.
4. Not enough passengers to pay for trip.
5. Overlapping start and end times.
6. Not a profitable route.
7. Bus not able to make trip (too many kilometers, too rugged).
8. Trip number is not unique.
9. Route segment already served by another bus.

Question 2:
A trip was scheduled by managers at Voyager Bus Inc. four months in advance of the start date. Yet when the time came for the bus trip to start, no bus showed up at the scheduled time. What reasons can you suggest for why there was no bus? (Provide as many solutions as you can think of.)

Answers
1. Bus driver was absent, not available (sick) - no replacement.
2. Mix-up in scheduling buses or drivers.
3. Bus broke down (not available).
4. Do details of schedule regarding the route segment to drivers or passengers.
5. Weather conditions or acts of God.

Question 3:
A bus driver for Voyager Bus Inc. has a problem. All seats on the bus have been taken, yet there is a passenger waiting to board the bus. What could have happened to cause this problem? (Provide as many solutions as you can think of.)

Answers
1. Passenger without reservation boarded the bus.
2. Head office does not keep track of how many passengers are assigned for each bus (overbooking).
3. Driver not aware of how many reserved seats are required.
4. Wrong bus was assigned to route.
5. Waiting passenger does not have reservation.
6. Some passenger boarded the wrong bus.
7. Waiting passenger has right route but wrong day (or wrong bus).
8. Bus has faulty seat.
9. One person takes up two seats.
10. Too many walk-up passengers allowed to board.

Question 4:
A person wants to go to a special event that is part of a trip offered by Voyager Bus Inc., yet the employee at Voyager Bus Inc states that, given the current situation, the person cannot go on the special trip. What could be stopping the person from going on the trip? (Provide as many solutions as you can think of.)

Answers
1. No more seats left on bus.
2. No buses are available (broken down).
3. Shortage of drivers (absent).
4. Passenger did not have reservation.
5. Voyager does not service the route segment.
6. Passenger has special needs that cannot be handled by Voyager (handicapped).
7. Trip requires special legal paper (Visa) or age restriction.
8. Employee is not correctly informed.
9. Event has been cancelled.
10. Part of the routes have been cancelled (bad weather, etc.).
11. Not enough demand for the trip.
12. Bus has already left.

Question 5:
Voyager Bus is considering the purchase of several new medium sized buses. What might be the effects of this purchase on Voyager Bus Inc. as it currently stands. (Provide as many solutions as you can think of.)

Answers
1. They can provide more trips.
2. Need to hire more drivers.
3. More maintenance problems.
4. Provide medium sized trips.
5. Serve more route segments (more new trips).
6. Decrease number of cancelled trips (more flexibility in schedule).
7. Where will the buses be stored?
8. Additional slack resources during off-peak season.
9. Retire older buses.
10. Less passengers need to reserve seats.
11. Drivers need different skills or licenses.
12. No attribute for bus size.
13. Record purchase of new bus.
14. Average number of customers per trip may be affected.
15. Lower maintenance since buses are new.
16. Record maintenance problems associated with new bus.

Problem Solving Questions and Answers: Far-Eastern Repair Facility

Question 1:
A customer of Far Eastern has called to complain that the machine they sent for repair has not been repaired yet. What possible reasons can you provide for what might have gone wrong. (Provide as many solutions as you can think of.)

Answers
1. No parts available (not yet available).
2. No mechanics available or assigned to repair.
3. No mechanic with skill available or assigned.
4. Part warehouse remote and difficult to access.
5. Machine already repaired but customer not contacted.
6. Machine already repaired but paperwork not complete or wrong.
7. May not be a machine type that Far East can repair.
8. Mechanics not able to solve problem.
9. No method for tracking machine in repair.
10. Duplicate value in repair number.
11. More repairs discovered.
12. Mechanics behind schedule.

Question 2:
Far Eastern is experiencing a very large increase in the number of machines that they should repair. What problems might Far Eastern experience because of this increase in repairs? (Provide as many solutions as you can think of.)

Answers
1. Not enough parts in inventory.
2. Not enough skilled mechanics.
3. Not enough mechanics.
4. Too many repairs assigned to mechanic.
5. More people to handle paperwork.
6. Need to efficiently schedule repairs.
7. Difficult to track machine being repaired.
8. Not enough space in shop or warehouse (backlog).
9. Backlog and delays of repairs.
10. Increased customer dissatisfaction.
11. Labor costs increase as mechanics may work more hours.
12. Not able to fix all machines.

Question 3:
Two customers of Far Eastern were talking to each other. One customer found that he was charged more for his repair than the other person, even though the machines were very similar. What reasons can you suggest for the difference in prices between the customers. (Provide as many solutions as you can think of.)

Answers
1. More mechanics were assigned to the repair.
2. Different parts were used.
3. Different mechanics may take different number of hours to complete the job.

4. Different mechanics get different labor rates and therefore different cost.
5. Machines states were different (older machine and more difficult to repair).
6. Repairs different although same machine.

Question 4:
Customers of Far Eastern are not happy when the actual repair price is higher than the estimated repair price. The sales person says that it is not his fault because the estimation is so difficult. Provide as many possibilities as you can think of that make the accurate estimation of the total repair price difficult. (Provide as many solutions as you can think of.)

Answers
1. Price of parts change (number of parts required).
2. Estimate of hours required by mechanics difficult.
3. Not sure who will do the repair.
4. Problems may be discovered after the repair begins.
5. Sales person not experienced with repair details.
6. State of equipment (age, disrepair) without taking apart (not sure what is wrong).
7. Availability of necessary equipment to complete the repair.

Question 5:
Far Eastern is considering investing in a machine that can be used to repair large turbine engines. How would the current data structure be affected by the purchase of the new machine? Try to think of as many affects as possible. (Provide as many solutions as you can think of.)

Answers
1. Nothing - any repair would still need a mechanic (update only).
2. Add a new classification for equipment that can be repaired.
3. Who will operate the machine? Need new mechanics, new skills.
4. Machine may not require mechanic to make repair. Add attributes to describe machine.
5. New warehouse for parts.
6. New attributes for equipment.
7. More data to be stored.
8. New suppliers and new parts added (attributes - part codes and bin number).
9. Repair details needs new attribute (machine used in repair).
10. Add entity called machine.
11. Include cost and cost estimates.
12. Cardinalities in some jobs may change.
13. Need to categorize size of engine for repair.

Appendix E: Cloze Test and Accepted Synonyms

Cloze Test: IFIP Conference Case

An IFIP Working Conference is an international conference centered on a topic of specific interest to one or more IFIP Working Groups. Participation in the conference is by _______1_______ only. Two objectives of the conference organizers are to ensure that members of the involved Working Group(s) and Technical Committee(s) are invited and that _______2_______is sufficient for financial break-even without _______3_______the capacity of the facilities available.

Two committees are involved in _______4_______an IFIP Working Conference: a Program Committee and an ________5______ Committee. The Program Committee deals with the _______6_______content of the _______7_______ and the _______8_______ Committee handles ________9______and local arrangements, along with _______10______and publicity. These committees ______11_______ together closely and have a need for common ______12_______.

When an IFIP conference is to be held, the Program Committee is responsible for several _______13______. These include preparing a _____14________ list of _______15______authors and sending a call for _______16______to each of these individuals. If the ______17_______ authors ______18_______with a letter of _______19______indicating they will submit a paper to the conference, the Program Committee ______20_______the participant's intent to participate. When the _______21______is received, the paper is also _____22________.
After receiving the papers, the Program Committee ________23_____a set of ______24_______for each paper and then sends the _____25________ to the respective _______26______. Reviewers create a _______27______report and send their reports to the committee. The committee then ______28_______accepted papers into groups and assigns session ______29_______. The list of _______30______ authors is forwarded to the ______31_______committee. This session information makes up the ______32_______for the conference.

The Organizing Committee begins by preparing a _____33________of people to invite to the conference. The committee issues ______34_______invitations to National ______35_______and to members of related ______36_______Groups. The organizing committee also _______37______that each of the contributing ______38_______ receive an invitation. Individuals who receive an _______39______and who intend to come to the conference must indicate their intent to _____40________by sending an ______41_______of invitation to the Organizing Committee. The final list of ______42_______is then generated by the ______43_______, who makes an effort to _______44______sending _______45______invitations to any individual.

Cloze Test: List of Accepted Synonyms

The following table lists the accepted synonyms for the Cloze test based in the **IFIP Conference** case as shown above. The number in the table corresponds to the number placed in the blank on the Cloze test above.

Blank No.	Correct Word	Synonym
1.	invitation	selection
2.	attendance	participation, number, amount, budget, revenue
3.	exceeding	overfilling, straining, surpassing, overcrowding,
4.	organizing	preparing, coordinating, planning, developing,
5.	organizing	N/A
6.	technical	content
7.	conference	N/A
8.	organizing	N/A
9.	financial	money, accounting
10.	invitations	N/A
11.	work	cooperate
12.	information	lists
13.	activities	tasks, functions, duties, responsibilities
14.	mailing	name
15.	potential	prospective, selected
16.	papers	submission
17.	potential	selected, prospective
18.	reply	answer, respond, accept
19.	intent	acceptance
20.	registers	files, records
21.	paper	submission
22.	registered	recorded, archived
23.	assigns	assembles, prepares, determines
24.	reviewers	N/A
25.	papers	submission
26.	reviewers	N/A
27.	review	N/A
28.	groups	organizes, collects. gathers, assigns
29.	chairs	heads
30.	accepted	N/A
31.	organizing	N/A
32.	itinerary	agenda, schedule
33.	list	mail-list
34.	priority	N/A
35.	representatives	N/A
36.	working	N/A
37.	ensures	requires, decides
38.	authors	members
39.	invitation	N/A
40.	participate	attend, come
41.	acceptance	confirmation, acknowledgment
42.	participants	attendees
43.	committee	organizers
44.	eliminate	prevent, stop
45.	duplicate	extra multiple, additional

Cloze Test: Entertainment Event Case

A company that owns a large sports and entertainment complex holds a variety of _____1_________ in their building including professional sports, concerts, and conventions. The _______2_______ of events is planned several months in advance. While no more than one event can take place at any given time, it is possible for more than one _____3_________ to occur on the same day.

To book an event, a _______4_______ contacts the ______5________ department for the complex and explains the type of event being promoted, the expected _______6_______, and a requested date and time for the event. After reviewing the _______7_______ application, the _______8_______ department either accepts the ______9________ and places the event in the _______10______ of events, or rejects the application. Once an event ______11_______ is accepted, the _______12______ is responsible for sending a ______13_______ plan for the event to the _______14______ department of the complex.

The operations department reviews all _______15______ plans and decides on the appropriate number of security personnel, _______16______, and customer service representatives for the ______17_______. These estimates for _______18______ requirements are _______19______ with the ______20_______ before they are finalized. Once the ______21_______ requirements are finalized, a _______22______ sheet is created by the operations department. The ______23_______ sheets lists the start time, end time, location, and _____24________ requirements for every _____25________ that will be required for the ______26_______. The completed ______27_______ sheets are then passed to the Human resource department who is responsible for _______28______ staff to work the event.

The human resource department creates a ____29_________ ________30______weeks in advance for all of the events to be held in the upcoming _____31______ week period. The ______32_______ sheets from all of the events in the _______33______ week period are used to assign _______34______ to the employees. The ______35_______ works through an entire deployment sheet one _______36______ at a time. ________37_____ are required to fill out an employee availability their ______38_______ and ________39_____. No employee can work over ______40_______ hours a week. In order for the _______41______ to assign a ______42_______ to an employee, the employee must have the right _______43______ and be available for the event. When the ______44_______ is finished this process, the final ________45_____ is posted for the employees.

Cloze Test: List of Accepted Synonyms

The following table lists the accepted synonyms for the Cloze test based in the **Entertainment Event** case. The number in the table corresponds to the number placed in the blank on the Cloze test above

Blank No.	Correct Word	Synonym
1.	events	entertainment, event
2.	program	N/A
3.	event	N/A
4.	promoter	N/A
5.	marketing	N/A
6.	attendance	seating, turnout, audience, people
7.	promoter's	promoters
8.	marketing	N/A
9.	application	proposal
10.	program	N/A
11.	application	proposal
12.	promoter	N/A
13.	seating	seat
14.	operations	operating
15.	seating	N/A
16.	cleaners	maintenance, cleaning
17.	event	N/A
18.	staffing	staff
19.	reviewed	discussed, consulted, confirmed, verified, checked
20.	promoter	N/A
21.	staffing	staff
22.	deployment	N/A
23.	deployment	N/A
24.	skill	N/A
25.	shift	duty, job, activity, deployment, position
26.	event	N/A
27.	deployment	N/A
28.	scheduling	HR, human resources
29.	schedule	N/A
30.	two	N/A
31.	two	N/A
32.	deployment	N/A
33.	two	N/A
34.	shifts	duties, jobs, activities
35.	scheduler	manager, department
36.	shift	job, duties, activities
37.	employees	workers, people
38.	availability	N/A
39.	skills	abilities, qualifications
40.	30	thirty
41.	scheduler	manager, department
42.	shift	job, duty, activity
43.	skills	abilities
44.	scheduler	manager, department
45.	schedule	N/A

Cloze Test: Voyager Bus Inc.

Voyager Bus Inc. (Voyager) is a company specializing in bus ______1________ to places of interest or special ______2________. The objectives of Voyager are to provide high quality and safe traveling experiences for _______3_______.

There are _______4_______ways for people to travel with Voyager. Passengers can either make a _______5_______ on a trip, or passengers can show up at the _______6_______ gate without a reservation and purchase a ticket for an ________7______ seat. Passengers with a reservation are assigned a _______8_______ date, whereas, passengers without reservations are assigned a ________9______ date. The name and addresses of ______10_______ passengers are collected. _______11______ numbers are collected where possible.

All bus trips are organized into daily route segments. _______12______ daily route segments have both a _______13_____ time and an end time. Each daily route segment Voyager organizes is _______14______ as a ______15_______ segment with a segment _______16_____, start town, and finish town. Voyager offers a range of trips, and each trip is made up of one or _____17________ route segments. For _______18______ trip there is a trip number, start town, and finish _______19______. If the trip is organized around a ______20_______ event, the event name is also associated with the______21_______.

______22_______daily route segment that Voyager offers is part of a daily _______23______. A daily trip is undertaken by one or more bus drivers. The name, address, and _______24______ number of all drivers is _______25______. Voyager also records information about ______26_______ drivers. When a driver is ______27_______, Voyager records the ______28_______ start date and the details about the _______29______. The ______30_______ driver provides one or more ______31_______ for being absent and each reason is assigned a _______32______l number and a short ______33_______.

Voyager also collects information about the ______34_______ used for daily trips. _______35______ have a make, model, and registration number. For buses in _______36______, the average daily kilometers is collected. If a bus requires ______37_______, Voyager notes the date on which the bus entered ______38_______ and records the one or more problems with the bus. Voyager assigns a ________39_____ number and a short ______40_______ for every _______41______ problem. Finally, the ______42_______ cost to repair all _______43______with a bus in _______44______ is also _______45______.

Cloze Test: List of Accepted Synonyms

The following table lists the accepted synonyms for the Cloze test based in the **Voyager Bus Inc.** case. The number in the table corresponds to the number placed in the blank on the Cloze test above

Blank No.	Correct Word	Synonym
1.	trips	tours
2.	events	N/A
3.	tourists	passengers
4.	two	2
5.	reservation	N/A
6.	boarding	departure
7.	unreserved	N/A
8.	reservation	N/A
9.	boarding	N/A
10.	all	every
11.	phone	telephone
12.	all	every
13.	start	beginning
14.	classified	organized
15.	route	trips, number
16.	number	N/A
17.	more	N/A
18.	each	every
19.	special	N/A
20.	town	city
21.	trip	tour
22.	each	every
23.	trip	tour
24.	employee	worker
25.	recorded	collected
26.	absent	sick
27.	absent	sick
28.	absence	sickness
29.	absence	sickness
30.	absent	sick
31.	reasons	excuses
32.	detail	N/A
33.	description	N/A
34.	buses	bus, vehicle
35.	buses	bus, vehicle
36.	use	operation, service
37.	maintenance	repair, repairs
38.	maintenance	N/A
39.	problem	N/A
40.	description	N/A
41.	maintenance	repair, repairs
42.	average	mean
43.	problem	N/A
44.	maintenance	N/A
45.	recorded	collected

Cloze Test: Far-Eastern Repair Facility

The Far-Eastern Repair Facility carries out _______1_______ of manufactured _______2_______ for clients. The objectives of the Far Eastern Repair Facility (Far Eastern) are to run an efficient repair facility that provides high quality repair service in the _______3_______ possible time. The repair facility has provided Far Eastern with the capability to repair _______4_______ general types of manufactured equipment: centrifugal pumps, reciprocating pumps, and diesel _______5_______.

When Far Eastern receives a piece of _______6_______ in need of _______7_______, the company assigns a ______8________ number and records the original equipment manufacturer's ______9________ along with the horsepower and _______10______ (RPM) at which the machine will run. For _______11______ pumps and diesel engines, Far Eastern also records both the piston_______12______ and cylinder volume associated with each piece of equipment. Every piece of equipment brought to Far-Eastern Repair is_______13______ by a _____14________. Far Eastern maintains the current address, and ______15_______ number of the customers for billing purposes. _______16______numbers are collected only for ______17_______ customers.

Far Eastern maintains an inventory of spare _______18______. The repair facility collects the ______19_______ and phone number of each______20_______ manufacturer so that they can_______21______ spare parts for machines they repair. The parts ______22_______is warehoused in three different buildings labeled warehouse 1, 2, and 3. Parts are stored in numbered ________23_____ within each warehouse. Each part is identified by a _______24______code along with the description, list price, ______25_______ number, and _______26______ for each part.

There are several _______27______at the facility. Since each mechanic differs in ______28_______ and experience, each mechanic has a different ______29______ rate. If a mechanic has a special _______30______, that ______31_______ is recorded by Far Eastern. Years of______32_______ is also recorded. When ______33_______ arrives for repair, one of the ______34_______ mechanics are ______35_______ to the repair task. The______36_______ mechanic then can ____37_________ other mechanics to _______38______ associated with the ____39_________. When the repair is completed, each mechanic working on the repair task enters the number of ______40_______ they spent on repair, parts_______41______ (if any), and a ______42_______ of the repair. The______43_______ cost associated with the ______44_______ task is then _______45______ and recorded.

Cloze Test: List of Accepted Synonyms

The following table lists the accepted synonyms for the Cloze test based in the **Far-Eastern Repair Facility** case. The number in the table corresponds to the number placed in the blank on the Cloze test above

Blank No.	Correct Word	Synonym
1.	repairs	repair, service
2.	equipment	engines, machines
3.	shortest	fastest
4.	three	3
5.	engine	engines
6.	machinery	equipment
7.	repairs	repair, service
8.	repair	N/A
9.	code	N/A
10.	speed	revolution
11.	centrifugal	N/A
12.	diameter	N/A
13.	owned	supplied, identified, coded
14.	customer	client
15.	phone	telephone
16.	fax	N/A
17.	local	N/A
18.	parts	part
19.	address	N/A
20.	parts	part
21.	order	purchase, buy, refill, restock, replace
22.	ordered	inventory, received, department, division, on hand
23.	bins	N/A
24.	parts	part
25.	bins	N/A
26.	weight	N/A
27.	mechanics	mechanic
28.	skill	skills
29.	labour	labor, wage
30.	skill	skills
31.	skill	skills
32.	experience	N/A
33.	equipment	engines, machines
34.	skilled	skills
35.	assigned	N/A
36.	skilled	skills
37.	assign	ask, appoint
38.	details	N/A
39.	repairs	repair
40.	hours	hour
41.	replaced	used, required
42.	description	N/A
43.	total	repair
44.	total	repair
45.	calculated	totaled, tallied, added, summed, invoiced, processed, charged

Appendix F: Ease of Use Questionnaire

Ease of Use Questionnaire

Entity Relationship Diagrams

I believe that it was easy for me to understand what an entity relationship diagram was trying to model.	0 - 7
Overall, I believe that the Entity Relationship Diagramming method was easy to use.	0 - 7
Learning how to read an Entity Relationship Diagram was easy for me.	0 - 7
Using the Entity Relationship Diagramming method was often frustrating.	0 - 7

Data Flow and Entity Relationship Diagrams

I believe that it was easy for me to understand what a data flow diagram together with an entity relationship diagram were trying to model.	0 - 7
Overall, I believe that the combination of a data flow diagram and an entity relationship diagram was easy to use.	0 - 7
Learning how to read a Data flow Diagram and an entity relationship diagram was easy for me.	0 - 7
Using the Data flow Diagram in combination with the entity relationship diagram was often frustrating.	0 - 7

Object Oriented Diagrams

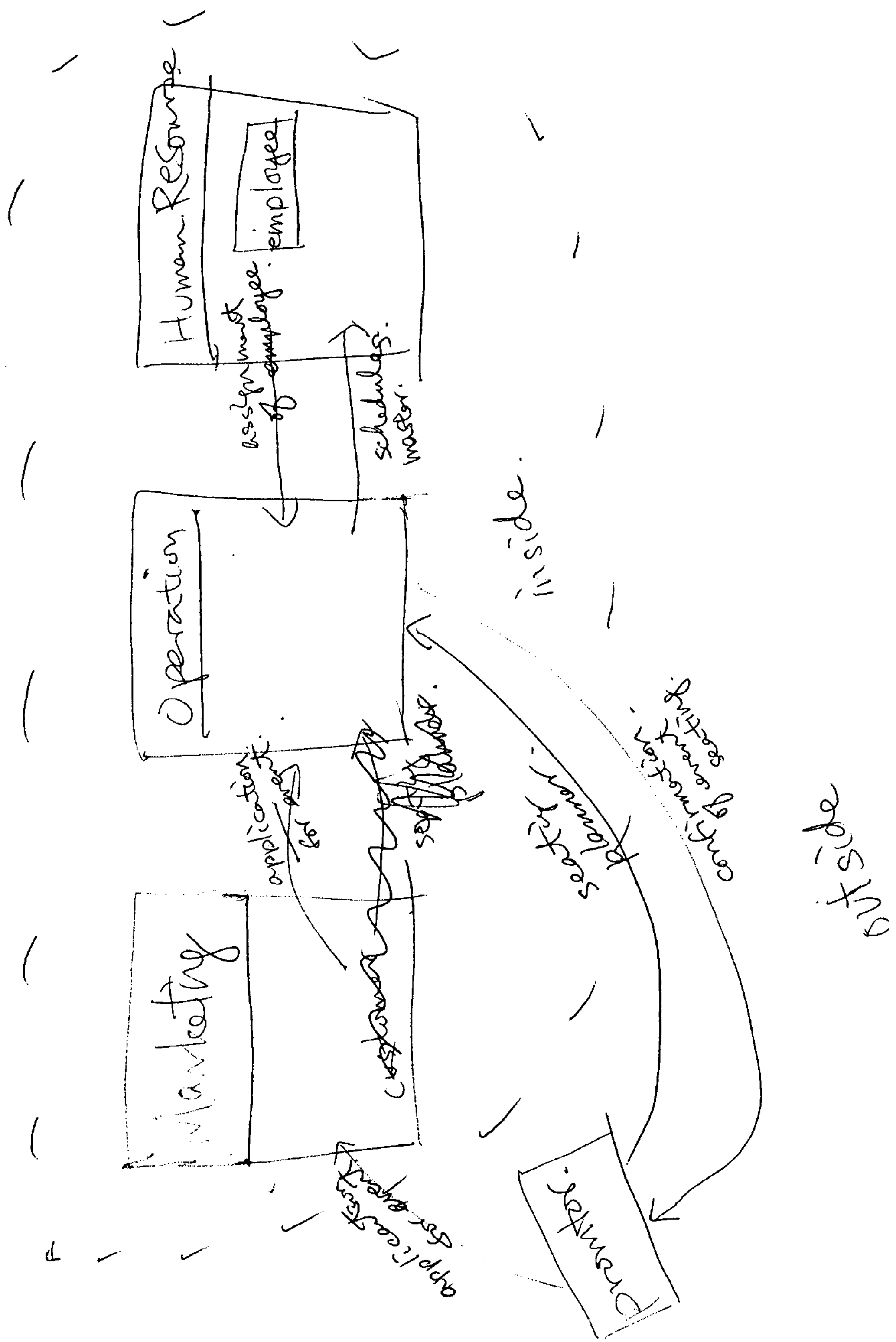
I believe that it was easy for me to understand what the object-oriented diagram was trying to model.	0 - 7
Overall, I believe that the Object Oriented Diagram was easy to use.	0 - 7
Learning how to read an Object Oriented Diagram was easy for me.	0 - 7
Using the Object Oriented Diagram was often frustrating.	0 - 7

Text (no diagrams)

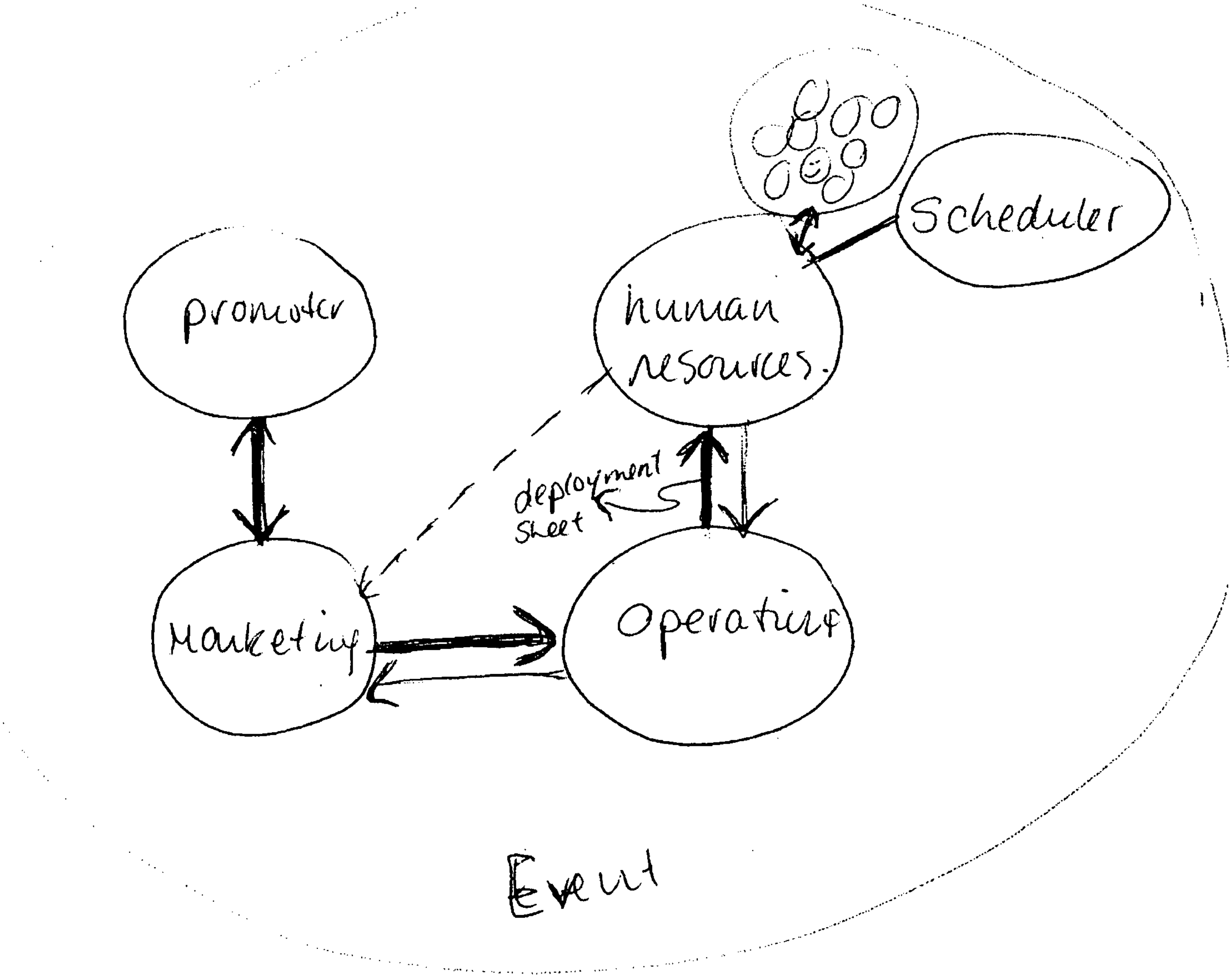
I believe that it was easy for me to understand what the text was trying to describe.	0 - 7
Overall, I believe that the text description is easy to use.	0 - 7
Learning how to interpret the text was easy for me.	0 - 7
Using the text was often frustrating.	0 - 7

Appendix G: Diagrams from Protocol Analysis

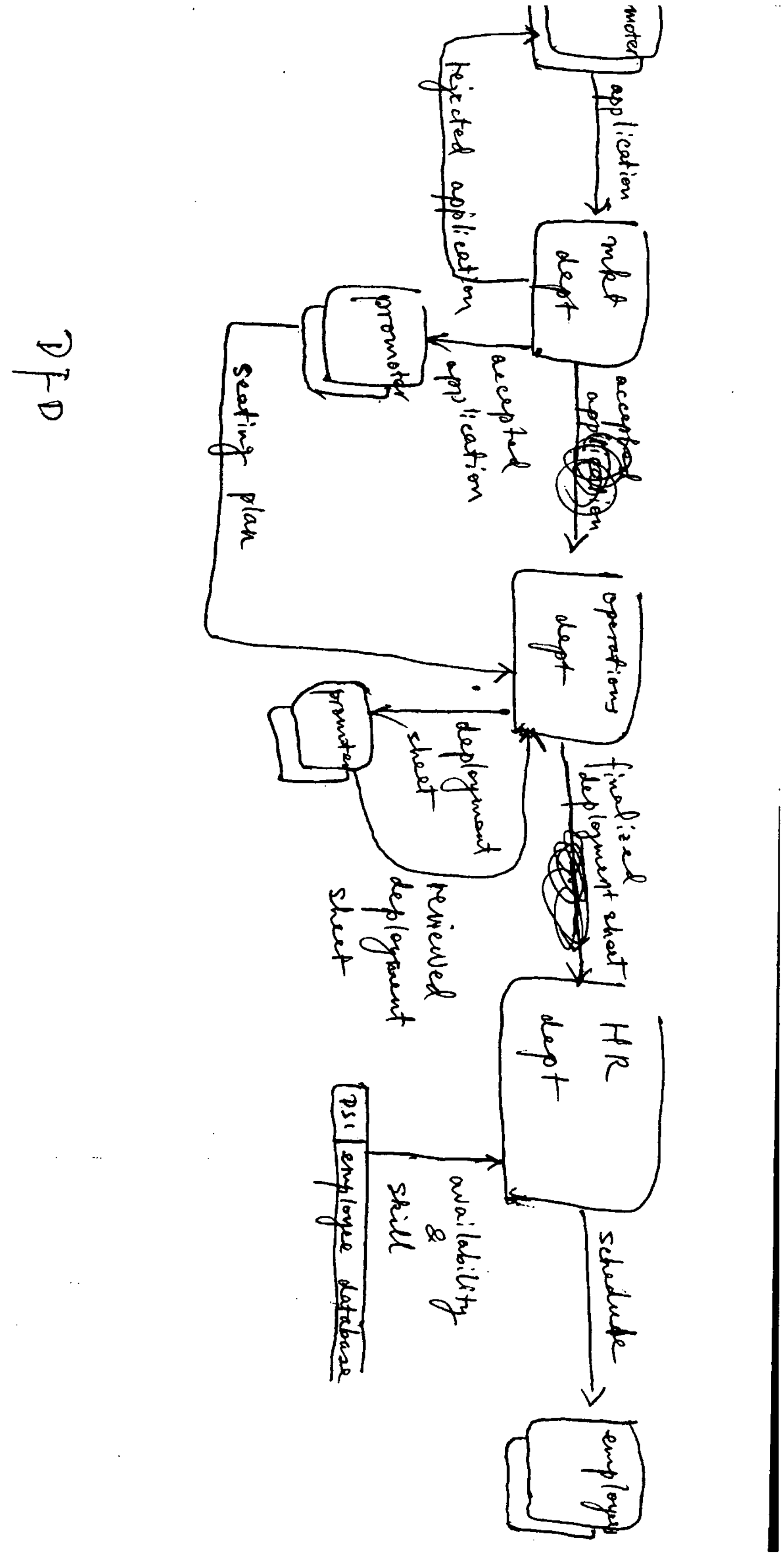
Participant 1: Text



Participant 2: Text



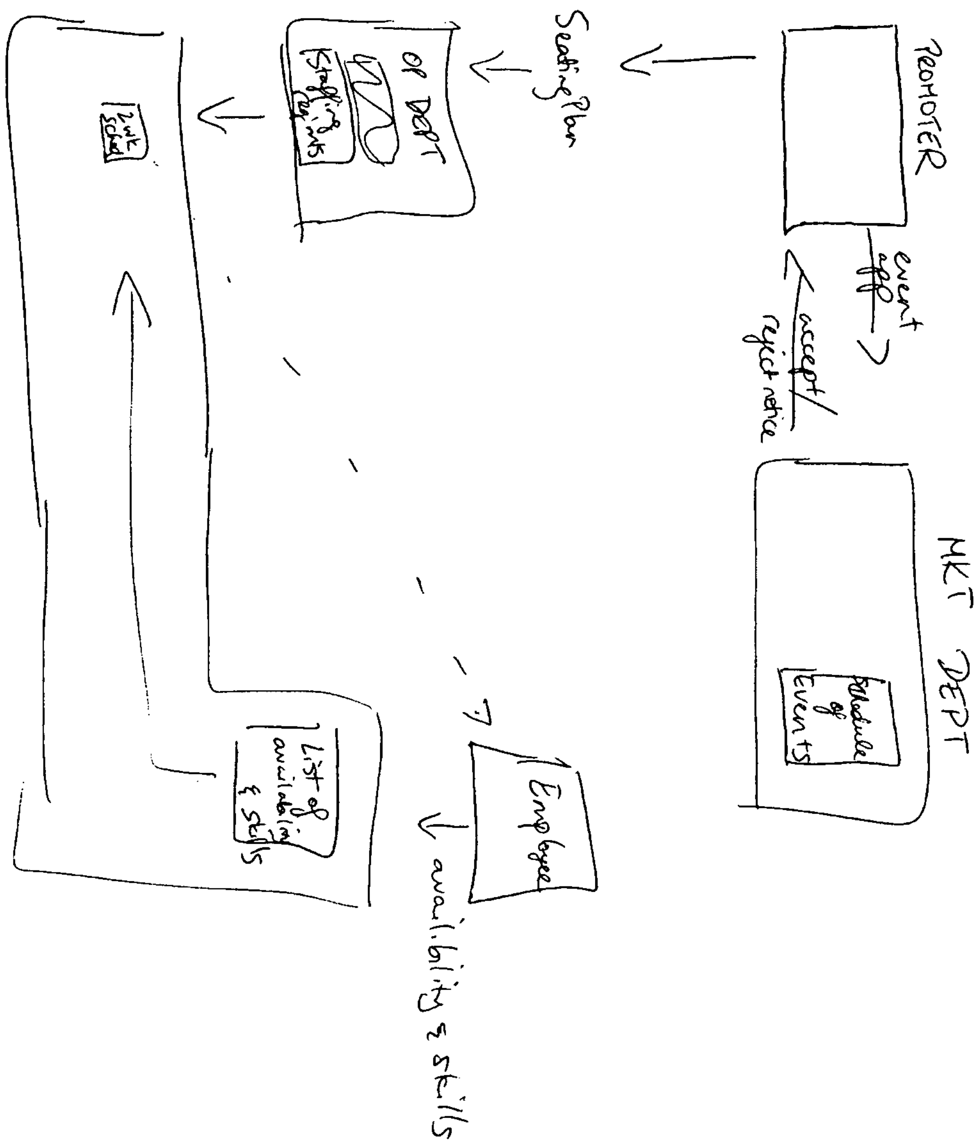
Participant 3: Text



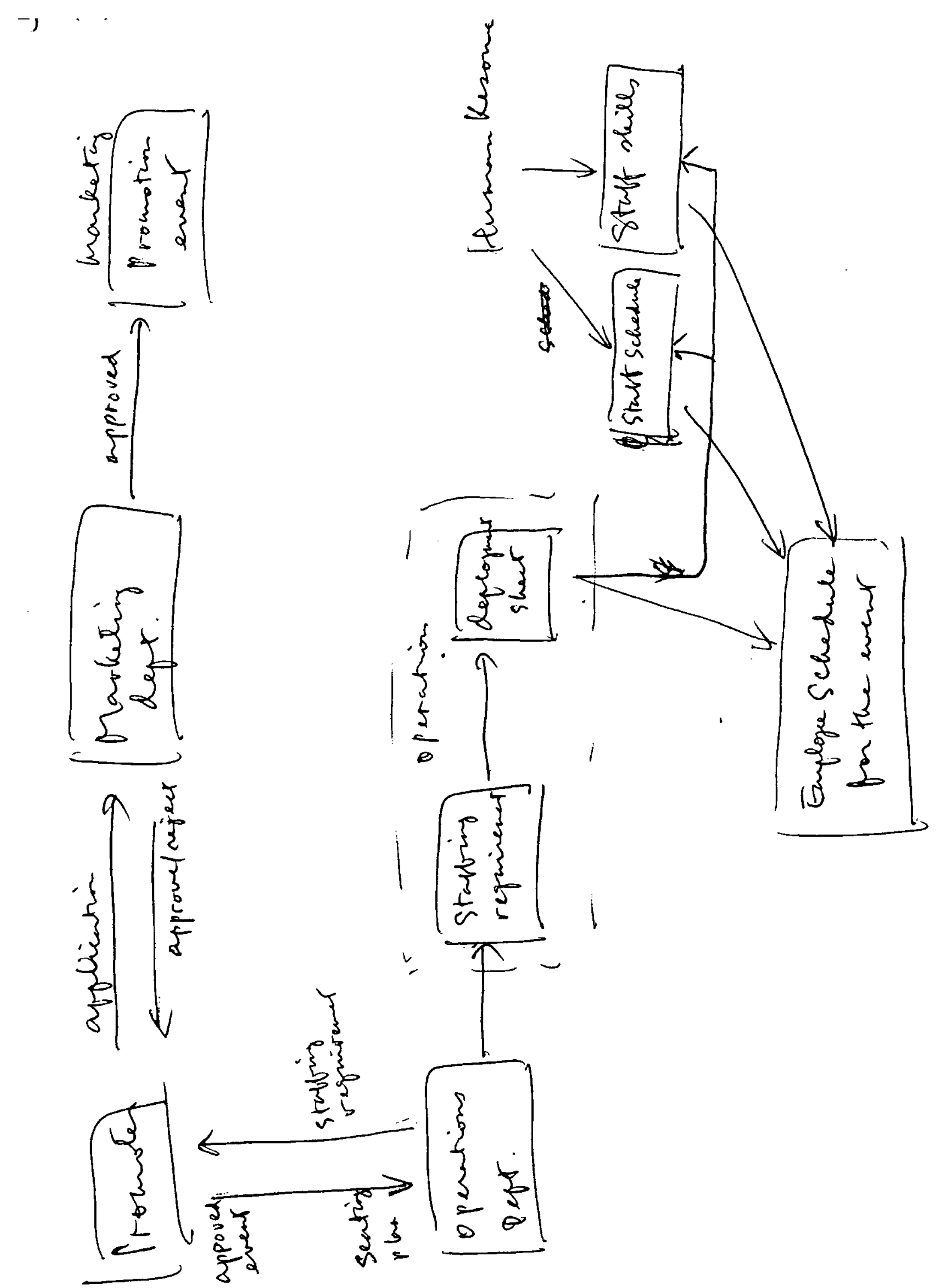
Participant 4: Text

1) Promoters Dept obtains/receives & application from Promoter
2) Promoters checks iternary of events
3) Promoters accepts/rejects application
4) Promoter sends in seating plan to Ops Dept.
5) Ops reviews seating plan & develops idea of supporting personnell
6) Ops & Promoter review staffing requirements
7) Ops created deployments sheets & transferred to the HR Dept
8) Employee availability cards & transfer them to the HR Dept.
9) HR depts schedules employees using the employee availability & deployment.
10) HR dept makes sure no one works 40 +
11) HR prepares a final schedule

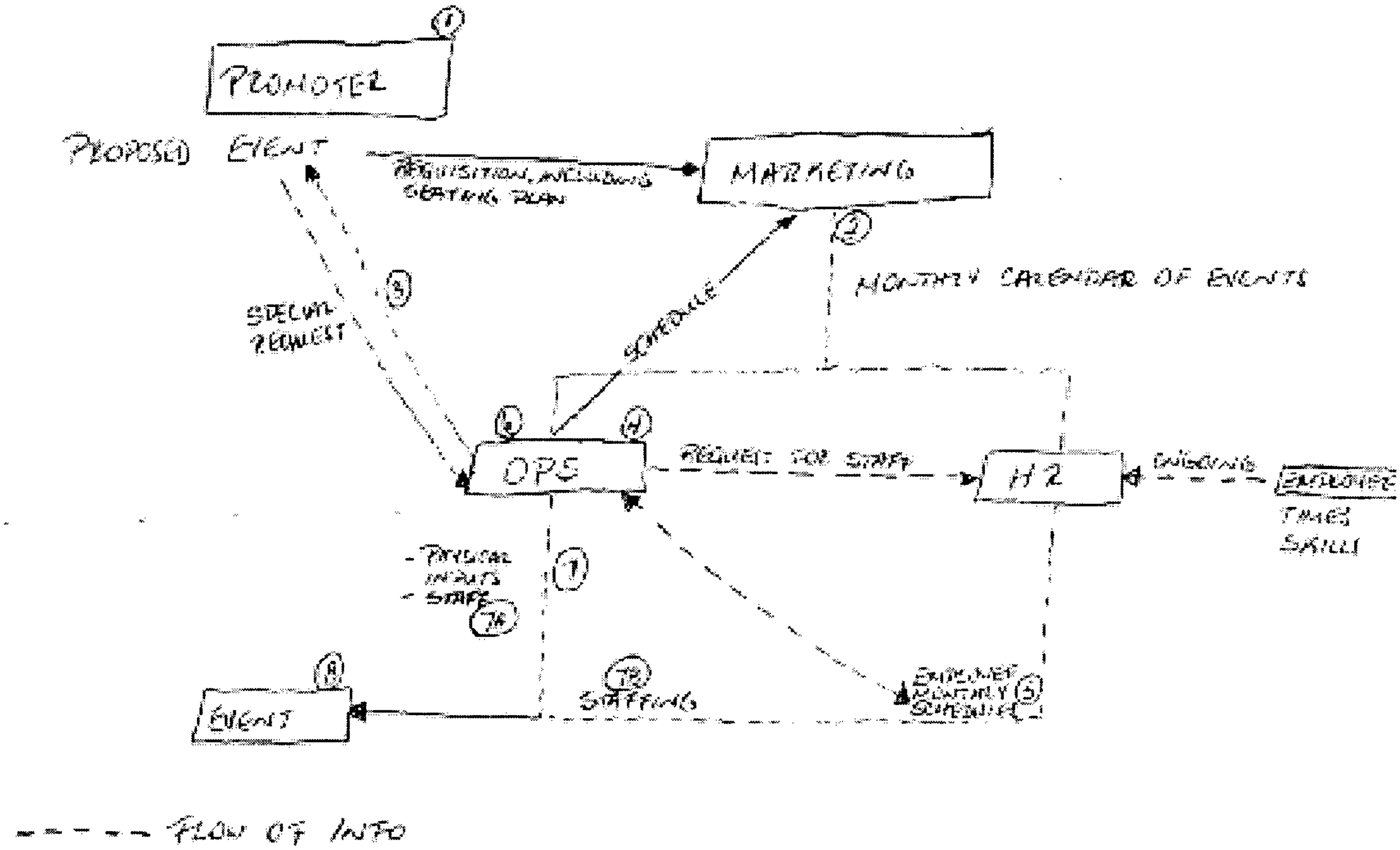
Participant 5: DFD and ER



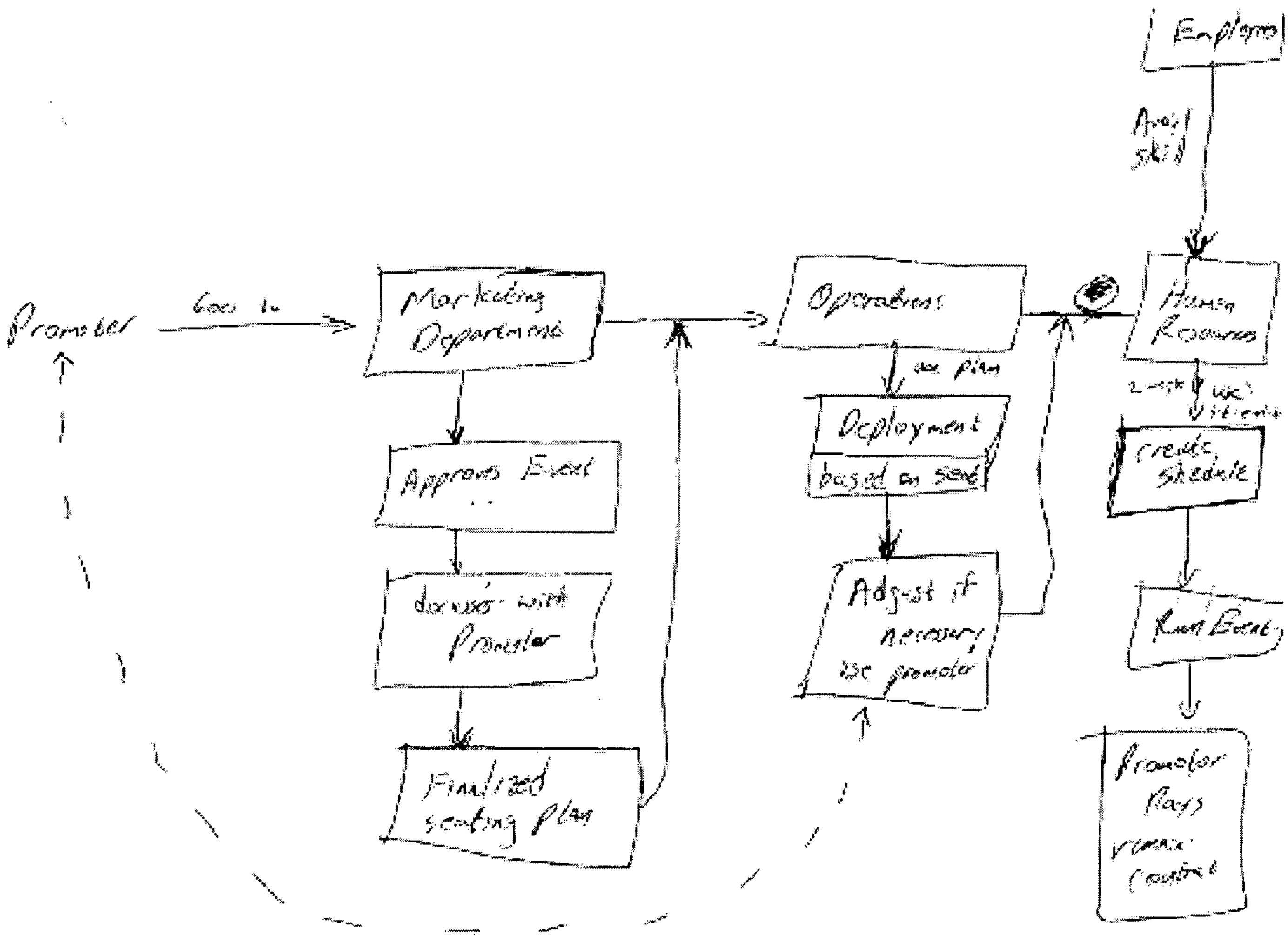
Participant 6: DFD and ER



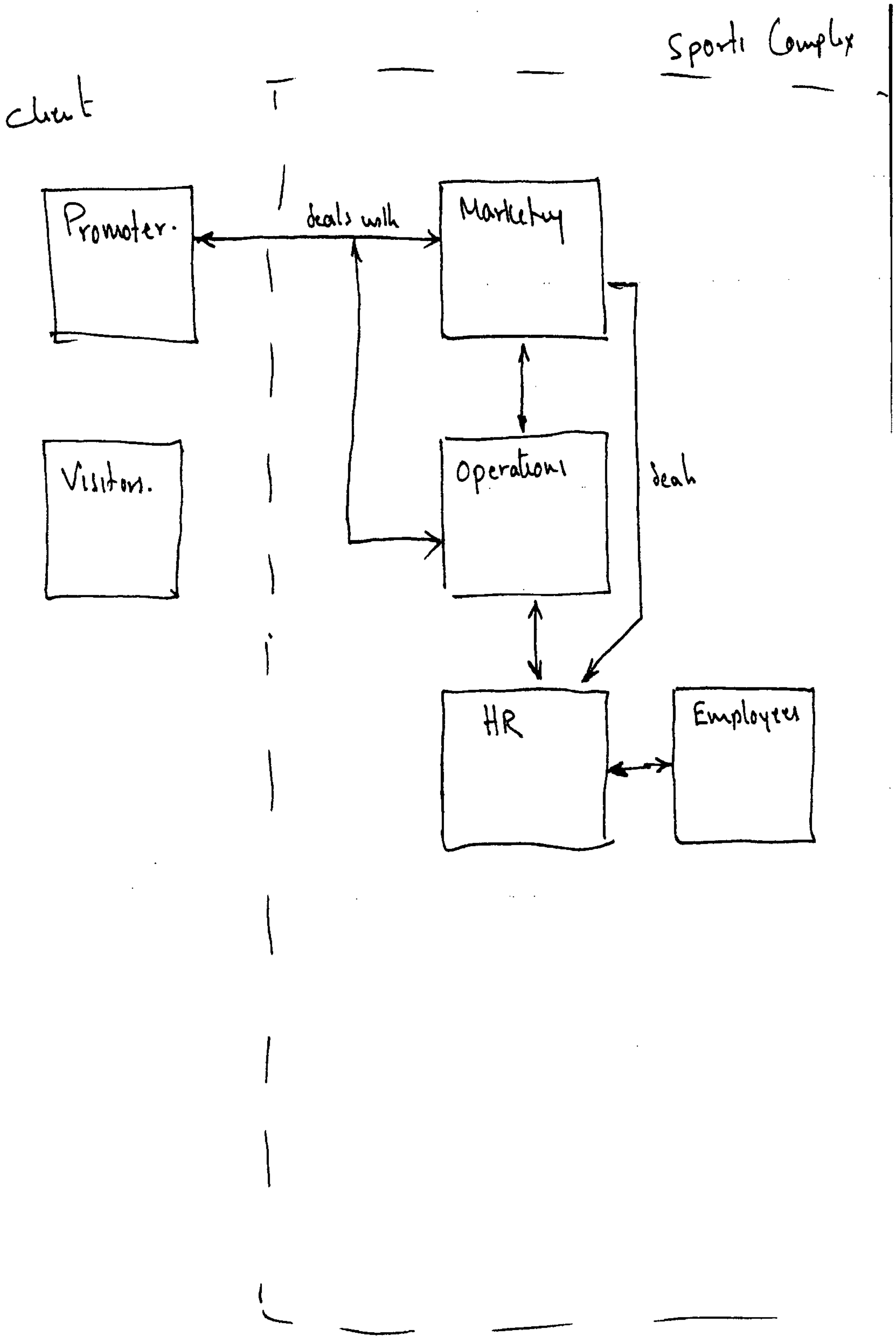
Participant 7: DFD and ER



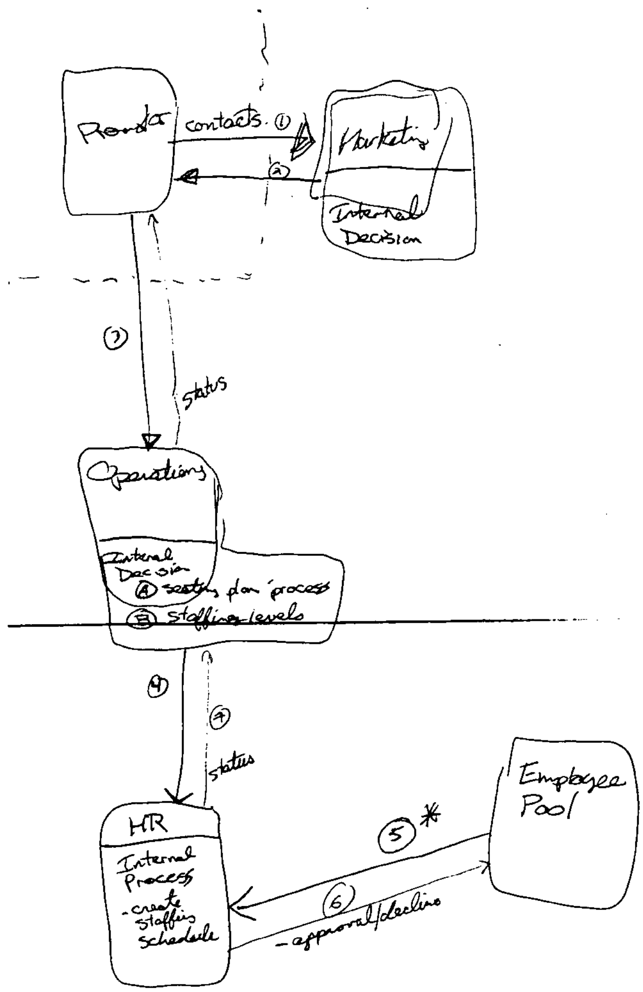
Participant 8: DFD and ER



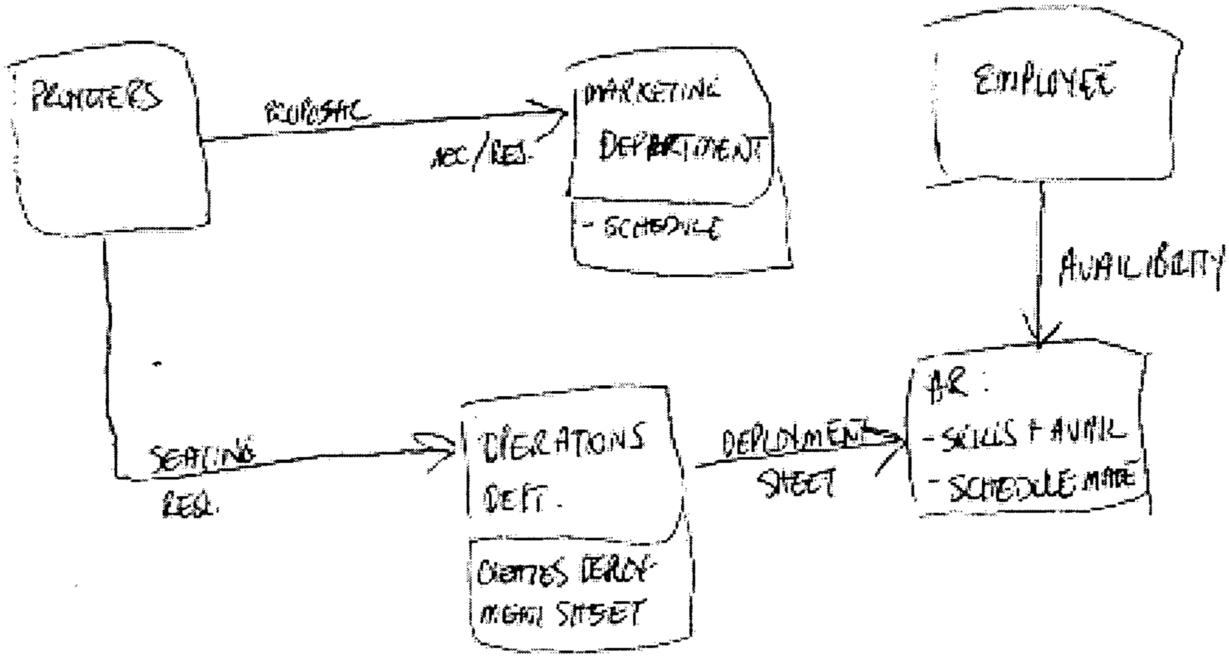
Participant 9: OOA



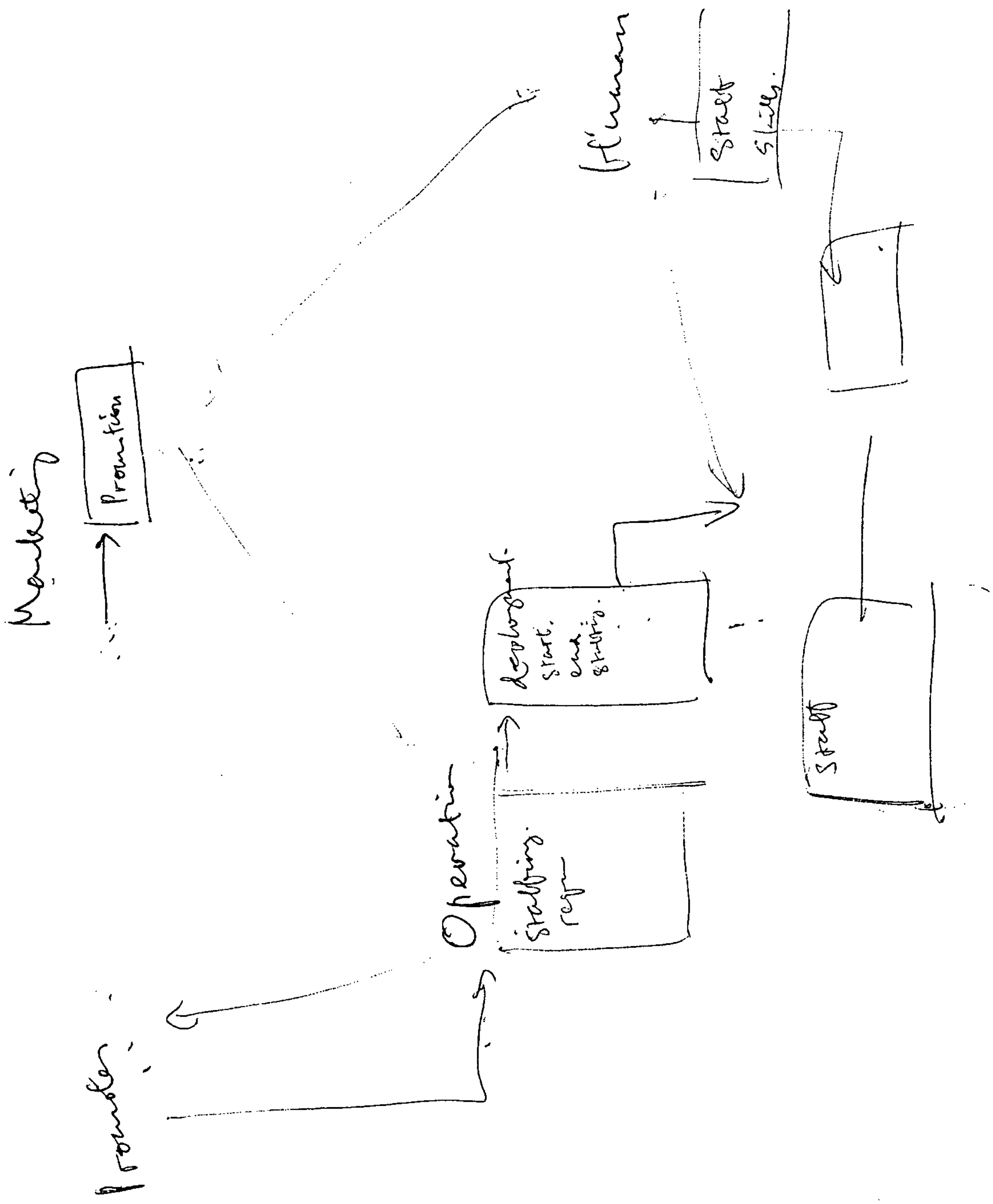
Participant 10: OOA



Participant 11: OOA



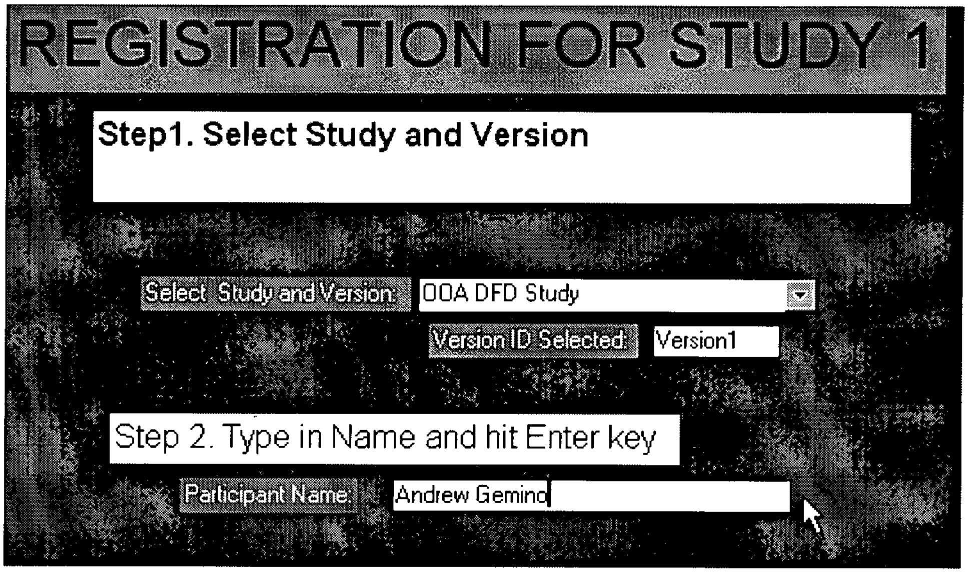
Participant 12: OOA



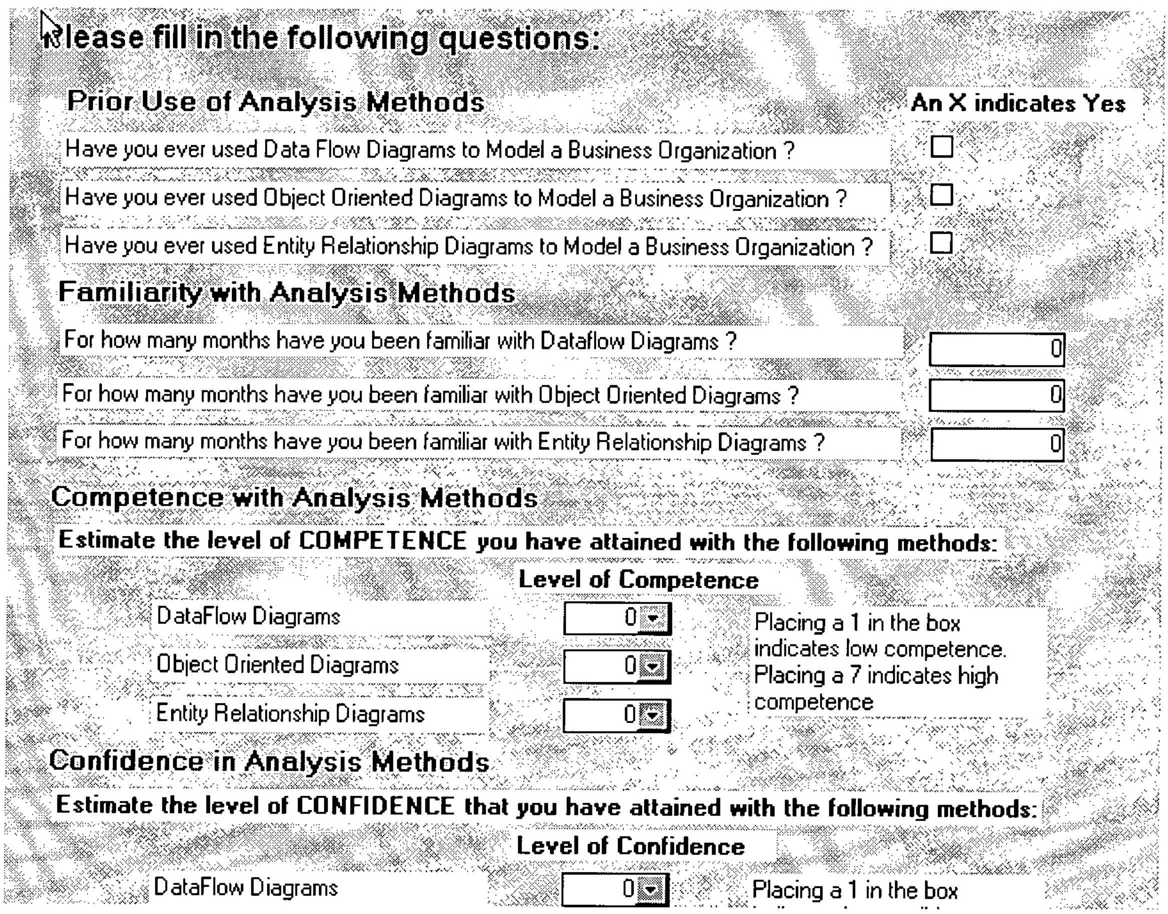
Appendix H: Screen Shots from Automated Entry Program

Screen Shots from Testing Program

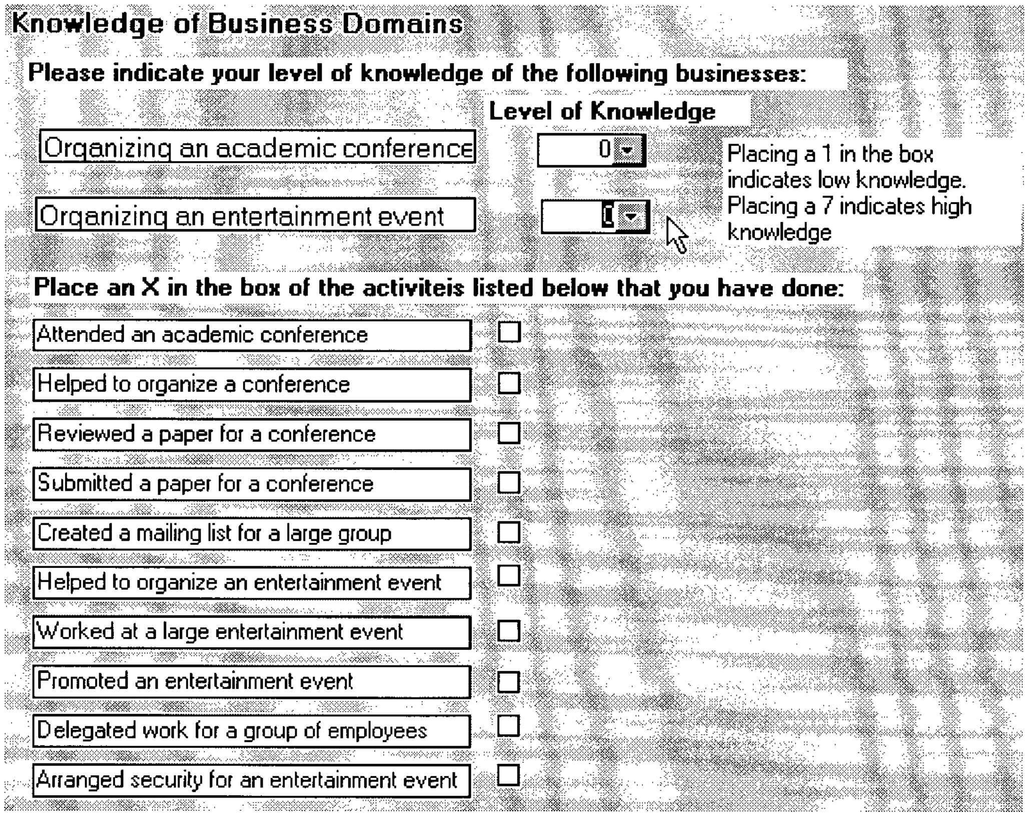
1. Registration Screen



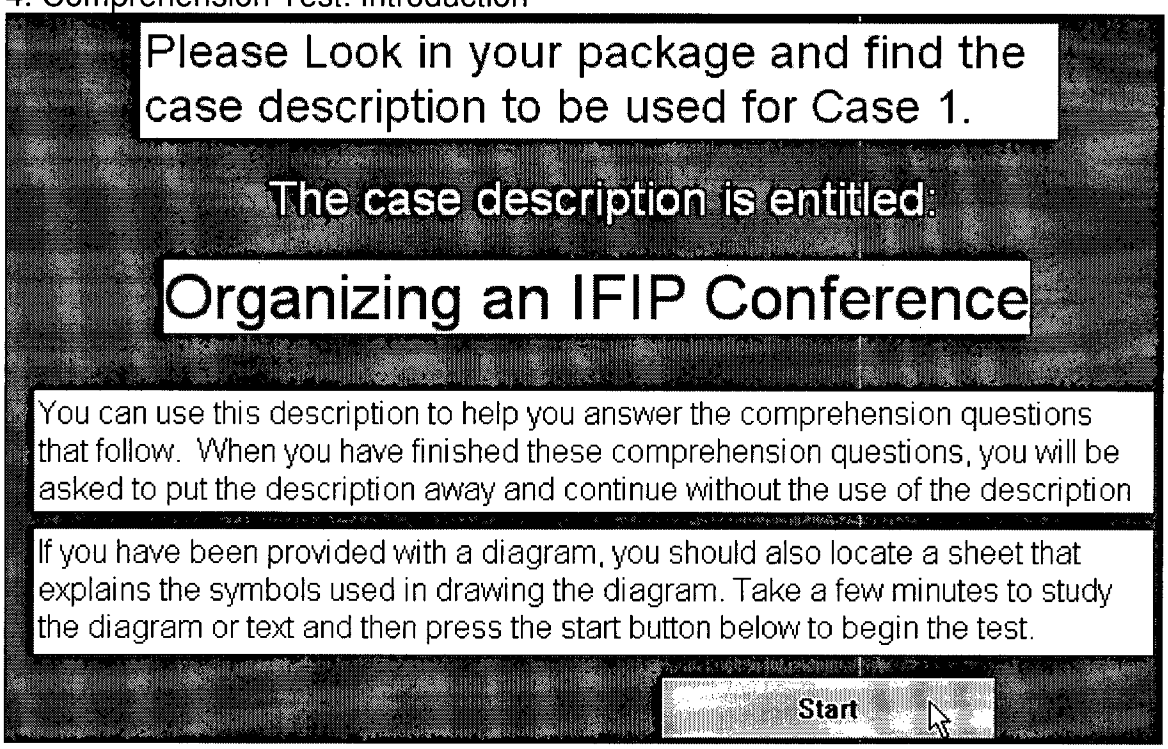
2. Pretest: Knowledge of Analysis Methods



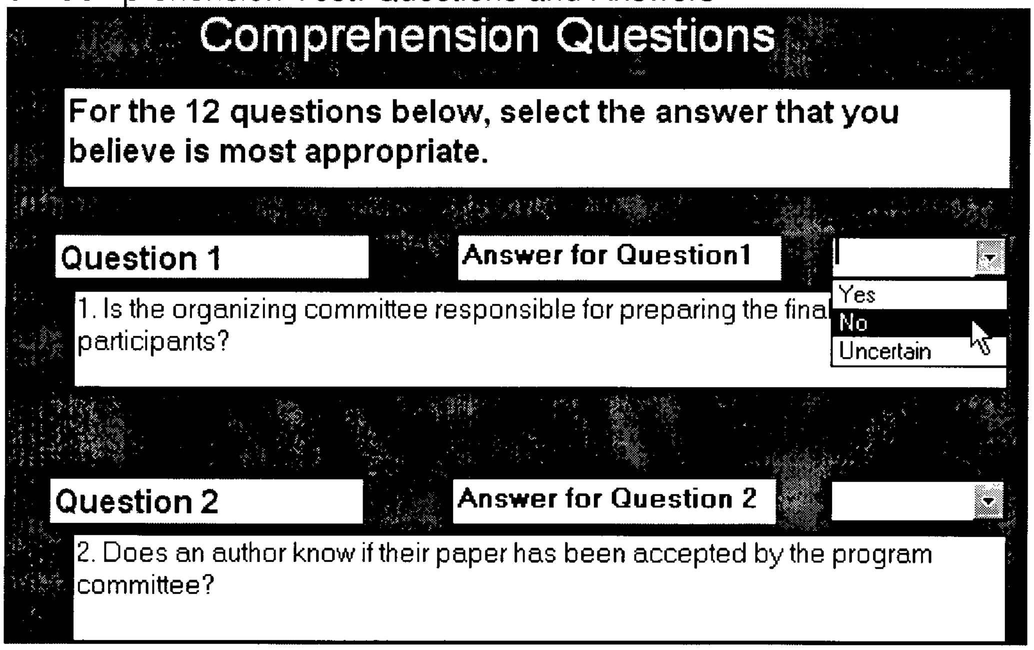
3. Pretest: Knowledge of Domain



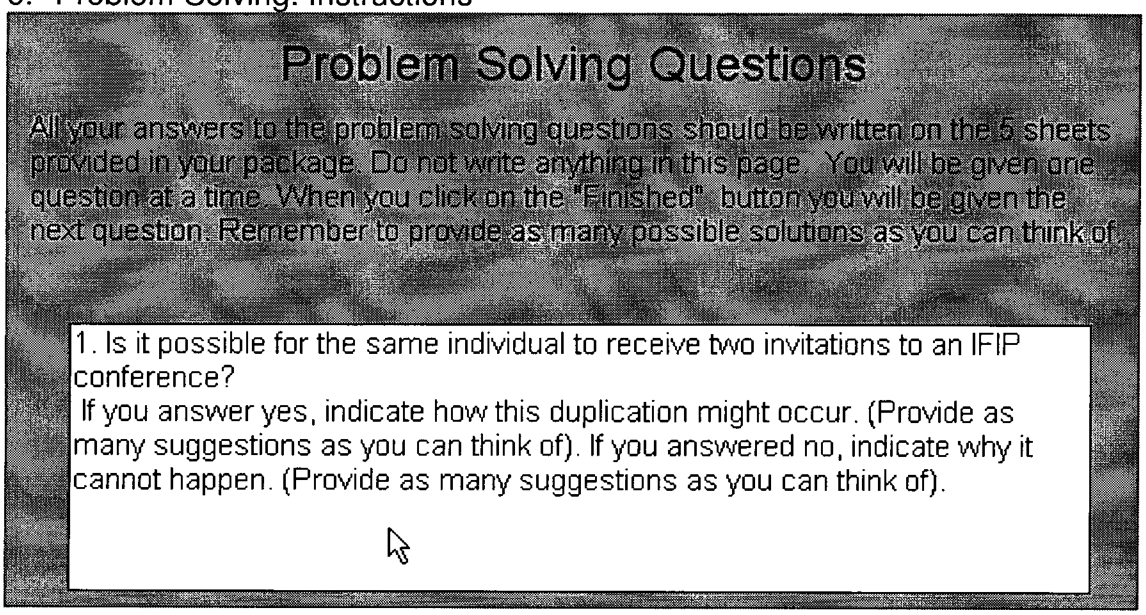
4. Comprehension Test: Introduction



5. Comprehension Test: Questions and Answers



6. Problem Solving: Instructions



7. Cloze Test: Introduction

Fill-in-the-Blanks Test

Now that you have finished the Problem Solving task, you have one more large task for this case. In this test you will be given a set of text that has been edited to remove some important words. Your job is to fill in the blanks left in the text using the knowlege you have gained from the description provided earlier.
Do NOT use the case description for this section of the test.

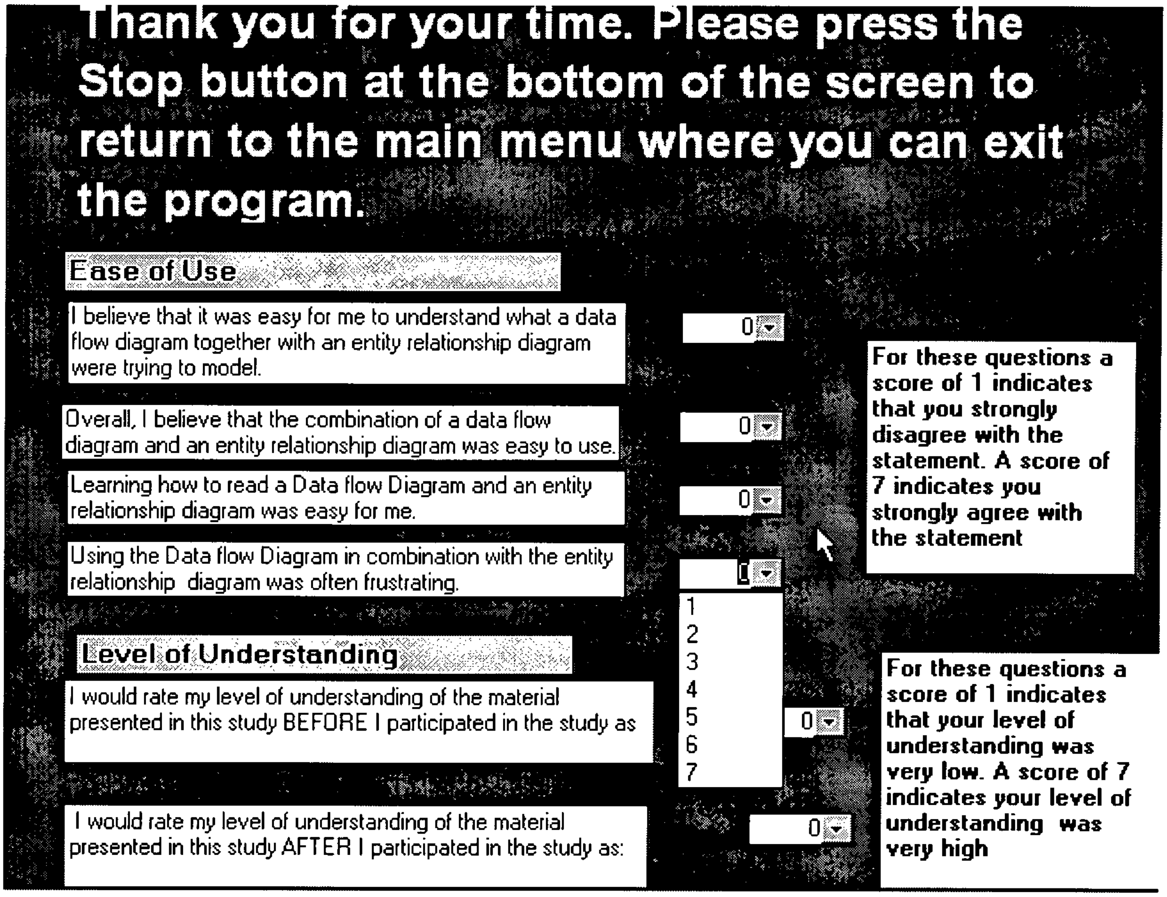
There is one thing to remember about this test. Each blank represents only a single word, so fill in the blank with the word you think best fits. When you are ready to start just press the button below.

8. Cloze Test

An IFIP Working Conference is an international conference centered on a topic of specific interest to one or more IFIP Working Groups. Participation in the conference is by ______ only. Two objectives of the conference organizers are to ensure that memebers of the involved Working Group(s) and Technical Committee(s) are invited and that ______ is sufficient for financial break-even without ______ the capacity of the facilities available.

Two committees are involved in ______ an IFIP Working Conference: a Program Committee and an ______ Committee. The Program Committee deals with the ______ content of the ______ and the ______ Committee handles ______ and local arrangements, along with ______ and publicity. These committees ______ together closely and have a need for common ______.

9. Ease of Use


Thank you for your time. Please press the Stop button at the bottom of the screen to return to the main menu where you can exit the program.
Ease of Use
I believe that it was easy for me to understand what a data flow diagram together with an entity relationship diagram were trying to model.
0
Overall, I believe that the combination of a data flow diagram and an entity relationship diagram was easy to use.
0
Learning how to read a Data flow Diagram and an entity relationship diagram was easy for me.
0
Using the Data flow Diagram in combination with the entity relationship diagram was often frustrating.
1
2
3
4
5
6
7
For these questions a score of 1 indicates that you strongly disagree with the statement. A score of 7 indicates you strongly agree with the statement
Level of Understanding
I would rate my level of understanding of the material presented in this study BEFORE I participated in the study as
0
I would rate my level of understanding of the material presented in this study AFTER I participated in the study as:
0
For these questions a score of 1 indicates that your level of understanding was very low. A score of 7 indicates your level of understanding was very high

Appendix I: Ontological Constructs in the Bunge-Wand-Weber Model

The following set of ontological constructs are provided originally by Wand & Weber (1995, p. 210-211). Wand & Weber have proposed that this set of constructs represents the minimum number of ontological constructs necessary to describe relevant items for system analysis and design.

Ontological Construct	Explanation
THING	A thing is the elementary unit in the BWW ontological model. The real world is made up of things. Two or more things(composite or simple) can be associated into a composite thing
PROPERTY INTRINSIC NON-BINDING MUTUAL BINDING MUTUAL EMERGENT HEREDITARY ATTRIBUTES	Things possess properties. A property is modeled via a function that maps the thing into some value. A property of a composite thing that belongs to a component thing is called an hereditary property. Otherwise it is called an emergent property. Some properties are inherent properties of pairs or many things. Such properties are called mutual. Non-binding mutual properties are those properties shared by two or more things that do not "make a difference" to the things involved; for example, order relations or equivalence relations. By contrast, binding mutual properties are those properties shared by two or more things that do "make a difference" to the things involved. Attributes are the names that we us to represent certain properties of things (normally abstract properties).
CLASS	A class is a set of things that can be defined via their possessing a single property.
KIND	A kind is a set of things that can be defined only via their possessing two or more properties.
STATE	The vector of values for all property functions of a thing is the state of the thing.
CONCEIVABLE STATE SPACE	The set of all states that the thing might ever assume is the conceivable state space of the thing.
STATE LAW: STABILITYCONDITION CORRECTIVE ACTION	A state law restricts the values of the properties of a thing to a subset that is deemed lawful because of natural law or human laws. The stability condition specified the states allowed by the state law. The corrective action specifies how the value of the property function must change to provide a state acceptable under the state law.
LAWFUL STATE SPACE	The lawful state space is the set of states of a thing that comply with the state laws of the thing. The lawful state space is usually a proper subset of the conceivable state space.
EVENT	An event is a change of state of a thing.
PROCESS	A process may be regarded either as an intrinsically ordered sequence of events on, or states of, a thing.
CONCEIVABLE EVENT SPACE	The event space of a thing is the set of all possible events that can occur in the thing.
TRANSFORMATION	A transformation is a mapping from one state to another state.

: (cont'd) Ontological constructs proposed by Wand and Weber	
LAWFUL TRANSFORMATION: STABILITYCONDITI ON CORRECTIVE CTION	A lawful transformation defines which events in a thing are lawful. The stability condition specifies the states that are allowable under the transformation law. The corrective action specifies how the values of the property functions(s) must change to provide a state acceptable under the transformation law.
EVENT SPACE	The lawful event space is the set of all events in a thing that are lawful.
HISTORY	The chronologically-ordered states that a thing traverses in time are the history of the thing.
ACTS ON	A thing acts on another thing if its existence affects the history of the other thing.
COUPLING: BINDING MUTUAL PROPERTY	Two things are said to be coupled (or interact) if one thing acts on the other. Furthermore those two things are said to share a binding mutual property (or relation); that is, they participate in a relation that "makes a difference" to the things.
SYSTEM	A set of things is a system if, for any bi-partitioning of the set, couplings exist among things in the two subsets.
Ontological Construct	Explanation
SYSTEM COMPOSITION	The things in the system are its composition.
SYSTEM ENVIRONMENT	Things that are not in the system but interact with things in the system are called the environment of the system.
SYSTEM STRUCTURE	The set of couplings that exist among things within the system, and among things in the environment of the system and things in the system is called the structure.
SUBSYSTEM	A subsystem is a system whose composition and structure are subsets of the composition and structure of another system.
SYSTEM DECOMPOSITION	A decomposition of a system is a set of subsystems such that every component in the system is either one of the subsystems in the decomposition or is included in the composition of one of the subsystems in the decomposition.
LEVEL STRUCTURE	A level structure defines a partial order over the subsystem in a decomposition to show which subsystems are components of other subsystems or the system itself.
EXTERNAL EVENT	An external event is an event that arises in a thing, subsystem, or system by virtue of the action of some thing in the environment on the thing, subsystem, or system.
STABLE STATE	A stable state is a state in which a thing, subsystem, or system will remain unless forced to change by virtue of the action of a thing in the environment (an external event).
UNSTABLE STATE	An unstable state is a state that will be changed into another state by virtue of the action of transformation in the system.
INTERNAL EVENT	An internal event is an event that arises in a thing, subsystem or system by virtue of lawful transformations in the thing, subsystem or system
WELL-DEFINED EVENT	A well-defined event is an event in which the subsequent state can always be predicted given that the prior state is known.
POORLY-DEFINED EVENT	A poorly-defined event is an event in which the subsequent state cannot be predicted given that the state is known.

Appendix J: An Overview of MANCOVA

Overview of Techniques Used in this Thesis

In an effort to reduce the material surrounding each of the many tests described in Chapter 6 and 7, the discussion of the background for the statistical techniques is summarized in this appendix.

The researcher makes use of three techniques for addressing the predictions and hypotheses listed in Table 5 of Chapter 4. These techniques are Analysis of Variance (ANOVA), Multivariate Analysis of Variance (MANOVA), and Multivariate Analysis of Covariance (MANCOVA). These three techniques share the same basic roots in the General Linear Model (GLM) that underlies many parametric statistical techniques. This section will briefly outline these three methods and the assumptions underlying the statistical techniques. This section is based largely on the discussion of ANOVA and MANOVA provided in Stevens (1992, Chapters 4, 5, and 6) and Hare et. al. (1992, Chapter 4).

ANOVA

ANOVA is used to compare the **means** of several populations, each of which is assumed to have the same variance. The comparison is based on a single **metric dependent** measure and one or more **independent categorical** variables. Equation (1) below displays the general form of the ANOVA model. The symbol Y_i indicates a metric dependent measure and X_i indicates a categorical independent variable:

(1) $$Y_1 = X_1 + X_2 + \ldots X_n$$

When a single categorical variable is used to in the ANOVA, it is referred to as a "One-way" ANOVA. Similarly, two categorical variables would signify a "Two-Way" ANOVA. Any ANOVA with more than a single categorical variable has the potential to observe **interaction** effects that may occur within different cells in the analysis. A cell indicates a potential combination of the categorical variables. For example, two categorical variables, each with two values will create four cells. Interaction effects occur when the differences across the groups created using one categorical variable are not distributed uniformly across all levels of the other categorical variable.

Assumptions Underlying ANOVA

The ANOVA technique is used to test the hypothesis that the means in cells, or groups of cells, are equal. To test this analysis we make use of an F statistic. The description of this test statistic is provided in Stevens (1992,p. 157). There are three assumptions underlying the development of the F statistic:

1. **The observations are independent (across all treatment groups)**
2. **The dependent variable is distributed normally.**
3. **The variances are equal for all treatment groups.**

In order to make valid inferences, in a formal sense, regarding populations from the F-test statistic, it is necessary to check that these three assumptions have not been violated. Nearly all samples taken from "real" populations will violate, to some degree, these assumptions. The question is whether the violations have a significant effect on the resulting F-test statistic. The popularity of ANOVA procedures is, in some part, due to the lack of sensitivity of the ANOVA F-test statistic to violations of these assumptions. As Hare (1992, p. 159) notes, "*There is evidence, however, that F tests in ANOVA are robust with regard to these assumptions except in extreme cases.*" A discussion of the three assumptions is provided below.

In regards to the first assumption of independence, Hare (1992, p. 160) notes that "*..there are no tests with an absolute certainty of detecting all forms of dependence.*" The best methods for avoiding problems with independence are good data collection procedures that do not create differences across treatment groups. Glass and Hopkins (1982, p.353) suggest,

> "*whenever the treatment is individually administered, observations are independent. But when treatments involve interactions between individuals... the observations may well influence each other.* "

Data collection procedures outlined earlier in Chapter 5 indicate that individuals were randomized into groups and that participants worked independently with the same set of materials under the same set of conditions. These procedures indicate that observations should be independent. The first assumption, at least theoretically, has been controlled for in the study design.

The second assumption in ANOVA is the normality of the dependent variable. Hare (1992, p. 160) notes that " *this (normality assumption) is the assumption for which violations have little impact.*" In a similar fashion, Stevens (1992, p. 238) notes "*that the type I error rate* *is essentially unaffected by non-normality. We say the F statistic is robust with respect to the normality assumption."* The normality assumptions are important, however, for interpreting the results of Bartlett-Box test statistic discussed below. For this reason, the normality of the dependent variables will be a considered in testing. As a first graphical test, stem and leaf plots for each of the dependent measures will be observed. There are several choices for numerical tests of normality. The Kolmogorov-Smirnov test statistic creates a combined consideration of both the skewness and kurtosis associated with a distribution. The null hypothesis is that the distribution is normal. Critical values for the test are provided in tables. The Wilks-Shapiro test for normality works in the same way and has been noted as being more sensitive to deviations from normal than the Kolmogorov-Smirnov test (Stevens 1992). A more complete procedure is to test the skewness and kurtosis separately. The reason for separate tests is provided by Steven (1992, p. 253) who notes that:

> *"We prefer using the skewness and kurtosis coefficients because we wish to separate out these two types of non-normality, whereas the Shapiro-Wilks statistic combines them. The reason we wish to separate them out is because kurtosis has been shown in both univariate and multivariate cases to have an effect on power, whereas skewness has been found not to effect power."*

In this study, we will consider all three of these tests when analyzing the normality of the dependent variable distributions.

The third assumption related to the ANOVA F-test is the homogeneity of variances. The Bartlett-Box homogeneity of variance test is used to determine whether variances are equal across all groups. The details of the test are provided in Stevens (Appendix A, p. 288-9). The Bartlett-Box test statistic is produced automatically in the ANOVA procedures with the Statistical Program for Social Sciences (SPSS ver 6.1)). This test, in conjunction with the normality of dependent variables will provide a test for homogeneous variances. Even if the tests show differences in the variances across groups, Stevens (1992, p. 241) notes the effect of heterogeneous variances will have ".*very slight effect on α (type I error) which is seldom distorted by more than a few hundredths."* Stevens (1992) also notes that the effects on type I error can be larger with unequal sample sizes. Fortunately, both studies in this thesis have relatively balanced designs.

This discussion has outlined the ANOVA procedure and has discussed the underlying assumptions and tests required to make inferences from the F-statistic produced by ANOVA. Fortunately, ANOVA is a robust technique that does not require strict adherence to the assumptions underlying the procedure in order to produce useful and valid results. In the next section, the ANOVA procedure is extended to include multiple dependent variables and the use of covariates in the MANOVA and MANCOVA procedures.

MANOVA and MANCOVA

The multivariate analysis of variance (MANOVA) technique extends the ANOVA technique described earlier by adding more dependent variables to the left-hand side of the equation (1). For a more complete description of the MANOVA techniques see Stevens (1992, Chap. 4, 5, 6, 9) or Hare (1992, Chap. 4). MANOVA is a more general form of the univariate analysis of variance (ANOVA). The general MANOVA model is displayed below in Equation (2). As in the ANOVA model, the symbol Y_i indicates a metric dependent measure, X_j indicates a categorical independent variable.

(2) $$Y_1 + Y_{2.} + Y_p = X_1 + X_2 + \ldots X_{nc}$$

The primary difference between MANOVA and ANOVA is the fact that the treatment groups are compared on *p* dependent variables simultaneously. In this way, the correlation between dependent measures is taken into account. Another difference is the test statistic used for hypothesis testing. The more familiar univariate analysis of variance (ANOVA), uses the F statistic for testing. The null hypothesis in these tests suggests that the means across treatment groups are equal. The test statistic most widely used for MANOVA, however, is Wilk's lamda (Λ). Wilk's Λ is similar in construction to the F statistic in that Wilk's Λ is a ratio of within group variability to the total variability from the grand mean for each variable. Wilk's Λ is a number, therefore, between zero and one. The larger the number, the greater the proportion of total variance explained by variation between groups.

The multivariate analysis of covariance (MANCOVA) extends the MANOVA procedure by including a set of metric independent variables on the right-hand side of equation (2) above. These metric independent variables are referred to as ***covariates***. The general model for the MANCOVA procedure is shown below. Note Z_i indicates the set of covariates.

(3) $$Y_1 + Y_{2.} + Y_p = X_1 + X_2 + \ldots X_n + Z_1 + Z_2 + \ldots Z_c$$

The primary difference between MANOVA and MANCOVA is that procedures similar to multiple regression are used to remove variation in the dependent variables due to the covariates. Once this variation is removed, a conventional MANOVA procedure is carried out on the adjusted variance for the dependent variables. Both the MANOVA and MANCOVA procedures are based on a set of assumptions. In any analysis, these assumptions must be addressed before analysis can begin. The assumptions for the MANOVA and MANCOA are discussed below.

Assumptions underlying MANOVA and MANCOVA

Our interests in the assumptions underlying the MANOVA are again directed by our interest in using the test statistics to infer differences between treatment groups. Meeting the mathematical assumptions underlying the MANOVA and MANCOVA analysis is necessary before making these inferences. The assumptions in MANOVA are somewhat more restrictive than in the ANOVA model. Stevens (1992, p. 237) notes that these restrictions need not distress researchers:

> *"in ANOVA and MANOVA we set up a mathematical model, based on these assumptions, and all mathematical models are an approximation to reality. Therefore, violations of the assumptions are inevitable. The salient question becomes, how radically must a given assumption be violated before it has serious effect on type I and type II error rates?"*

This section will provide only a cursory overview of the assumptions and the effects of their violation. For a more complete discussion see Stevens (1992, Chapter 6, 9).

There are three assumptions underlying the MANOVA procedure. Three additional assumptions underlie the MANCOVA procedure. The MANOVA assumptions are discussed below followed by the MANCOVA assumption. The first three assumptions are similar to those in the ANOVA model:

1. **The observations are independent (across all treatment groups).**
2. **The observations on the p dependent variables follow a multivariate normal distribution.**
3. **The population covariance matrices for the p dependent variables in each group are equal**

Our previous discussion of the assumptions indicated that the first assumption, independence, was an important assumption, and that violation of this assumption has large effect on the level of type I error. We have also discussed the fact that proper experimental design should ensure independence between observations. A close look at preliminary statistics in both studies will help to ensure that there are no obvious violations of the first assumption. This is done later in the chapter.

The second assumption of multivariate normal populations is more restrictive than ANOVA as normality on each of the dependent variables is a necessary but not sufficient condition. This means that even when all dependent variable populations are normal, the resulting multivariate distribution may not be normal. Thankfully, Stevens (1992, p. 247) notes that for MANOVA *"deviation from multivariate normality has only a small effect on type I error."* Assessing the multivariate normality can be difficult as Stevens (1992, p. 248) notes, *"Unfortunately, none of the major statistical packages (SPSS, BDMP, SAS) have a test of multivariate normality."* While there is no direct test for multivariate normality, we can rely on the same univariate normality measures to at least verify normality for each dependent variable. While these univariate tests are not sufficient for determining multivariate normality, they provide at the minimum an indication of the potential for multivariate normality.

The final assumption in MANOVA is that the covariance matrices for the dependent variables are equal. As Stevens (1992) noted, it is unlikely that this assumption would ever literally hold. Fortunately, when group sizes are approximately equal the actual level of type I error is very close to the estimated type I error. There is a small effect on type I error when the covariance matrices are not homogenous. Unlike the multivariate normal assumption, a test for the homogeneity of the covariance matrix does exist. The test for homogeneous variance is called the Box's M test and is described in Stevens (1992, p. 268-9). The Box's M test is created in SPSS under the MANOVA option. Two statistics are provided; the F–test, and the Chi-square test. Stevens (1992, p. 260) provides some guidance in choosing between the test statistics:

> *"When all group sizes are greater than 20 and the number of dependent variables is less than 6 and the number of groups is less than 6 than the chi-square approximation is fine. Otherwise the F approximation is more accurate and should be used"*

Since the first three conditions in the above quote are met in this thesis, the chi square statistic for the Box's M test will be used.

As noted earlier, three additional assumptions must also be considered when applying MANCOVA. These assumptions are:

4. **A linear relationship exists between the dependent and covariate variables.**
5. **For two covariates (as used in this study), the regression planes must be parallel across treatment groups.**
6. **The covariates should be measured without error.**

The restrictions placed on MANCOVA by these assumptions is relatively severe and even worse, the violations of these last three assumptions are serious as they can lead to inaccurate differences reported between treatment groups.

The fourth assumption, linear relations between dependent and variables and covariates, can be tested with a simple correlation analysis. This assumption suggests that variables that are not linearly related to the dependent variables should not be included in the final model. If the relationship is curvilinear, there are transformations that are possible, or the model could be estimated using curvilinear, but these are rarely used (Stevens;1992, p. 334).

The requirement for parallel regression planes suggests that there ***is no covariate by treatment interaction effect***. This assumption guarantees that a difference between populations that results from the current sample would be observed in other samples because the difference between groups is constant (the regression lines are parallel) for all observed levels of the sample mean. To test this assumption, we set up a covariate by treatment group effect in the MANCOVA test procedure and test to see if the interaction effect is significant. If it is significant, then the fifth assumption is violated and the level of type I error could be affected substantially. This violation is serious and would require attention.

The sixth assumption, that the covariates be measured without error, is a difficult assumption to prove, but very important to consider. Steven's notes that:

> *"There is always measurement error for the variables that are typically used as covariates in social science research. And measurement error causes problems in both randomized and non-randomized designs. But it is more serious in non-randomized designs... This problem would not be of particular concern if we had a very reliable covariate like I.Q. or other cognitive variables from good standardized tests. If, on the other hand, the covariate is a noncognitive variable, or a variable*

derived from a nonstandardized instrument, then concern would definitely be justified."

The violation of this assumption, as noted above, is more serious in nonrandomized designs. Stevens (1992, p. 334-5) notes that:

" In the case of randomized designs.... the power of the ANCOVA is reduced to what it would be if no error were present, but treatment effects are not biased. With other designs the effects of measurement error in covariates are likely to be serious."

For randomized designs, therefore, measurement error reduces the power of the test, (increases the likelihood of type II error), but does not have a serious effect on the level of type I error.

The discussion above suggests that we look closely at the covariates that we will use for our tests. In randomized designs, the more accurately and consistently the covariate can be measured, the lower the level of type II error that we accept. The discussion also suggests that the randomized design described in the earlier chapter, help to protect the study from potentially serious effects on treatment effects. Finally, analysis using more powerful statistical techniques such as MANOVA and MANCOVA is made easier when independence and equal group sizes are present in a design.

In this section, we have described the MANOVA and MANCOVA procedures and the assumptions underlying these techniques. We have also outlined the effects that can occur when the assumptions are violated and how to test for the violation of assumptions. This information is summarized in Table 10 below.

Assumptions and their Effects on MANOVA and MANCOVA

Assumption	Effect of Violation	Test for Violation	Controlling against Violation
1. The observations are independent (across all treatment groups).	Serious, type 1 error affected, bias of treatment effect	No standard tests, observation	Randomize testing
2. The observations on the p dependent variables follow a multivariate normal distribution	Not serious. Small effects on type I error except in extreme cases	No direct test for multivariate. Use univariate tests (Kolmogorov-Smirnov, Shapiro-Wilks, and or Kurtosis/Skewness)	No control available. Variables can be transformed if necessary
3. The population covariance matrices for the p dependent variables in each group are equal	Not serious. Small effects on type I error except in extreme cases	For multivariate, the Box test for homogeneity of variance	Large sample size
4. A linear relationship exists between the dependent and covariates..	Serious bias of treatment affect.	Pearson's Correlation between covariates and dependents	Standardize pre test questions
5. For two covariates, the regression planes must be parallel across treatment groups	Serious bias of treatment effect. Affect on type ! could be large	Test for covariate by treatment interaction by creating a contrast. There should be no significant effect	Randomize subjects across groups.
6. The covariates should be measured without error.	In random designs affects type II error	No standard tests	Standardized, well developed measures