8-22-2007

The Model-Based Systematic Development of LOGIS Online Graphing Instructional Simulator

Darrel R. Davis
University of South Florida

Scholar Commons Citation

Davis, Darrel R., "The Model-Based Systematic Development of LOGIS Online Graphing Instructional Simulator" (2007). *Graduate Theses and Dissertations.*
https://scholarcommons.usf.edu/etd/690

The Model-Based Systematic Development of LOGIS Online Graphing Instructional Simulator

by

Darrel R. Davis

A dissertation submitted in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
Department of Secondary Education
College of Education
University of South Florida

Major Professor: James White, Ph.D.
Darrel Bostow, Ph.D.
Robert Dedrick, Ph.D.
William Kealy, Ph.D.

Date of Approval:
August 22, 2007

Keywords: Developmental Research, Simulation, ADDIE, Model, Systematic Design, Guided Contingent Practice, Programmed Instruction, Adaptive Instruction

Dedication

To Martha, Ronald, June, Ray, Sharret, and Shawn

Acknowledgements

I would like to thank the members of my doctoral committee for their support throughout this dissertation process. To Dr. White and Dr. Bostow, you have been tremendous! Thank you both for taking the journey with me. I could not have asked for better mentors. To Dr. Dedrick, your detailed analyses were invaluable. To Dr. Kealy, thank you for your support.

To my parents June and Ray, I am because of you. To my wife Martha, "you are…" To my brother Ronald, thank you for believing in me. To Shawn, thank you for being there for me. To Ms. Ellie, thank you for everything.

To the hidden gems in life who have been a constant source of encouragement. To Imed, Ilir, Karim, Jeff, Mike, Kale, Carol, Kelly Gummi, Satch, and Mrs. Hendrix, you have all been incredible gems.

TABLE OF CONTENTS

Chapter Three Proposed Development 85

LIST OF TABLES

LIST OF FIGURES

The Model-Based Systematic Development of Logis
Online Graphing Instructional Simulator

Darrel R. Davis

ABSTRACT

This Developmental Research study described the development of an interactive online graphing instructional application and the impact of the Analysis Design Development Implementation Evaluation (ADDIE) model on the development process. An optimal learning environment was produced by combining Programmed Instruction and Adaptive Instruction principles with a graphing simulator that implemented guided contingent practice. The development process entailed the creation and validation of three instruments measuring knowledge, skills, and attitudes, which were components of the instruction.

The research questions were focused on the influence of the ADDIE model on the development process and the value of the LOGIS instructional application. The model had a significant effect on the development process and the effects were categorized by: Organization, Time, and Perspective. In terms of Organization, the model forced a high level of planning to occur and dictated the task sequence thereby reducing frustration. The model facilitated the definition of terminal states and made it easier to transition from completed tasks to new tasks. The model also forced the simultaneous consideration of global and local views of the development process.

The model had a significant effect on Time and Perspective. With respect to Time, using the model resulted in increased development time. Perspectives were influenced because previously held assumptions about instructional design were exposed for critique. Also, the model facilitated post project reflection and problem diagnosis.

LOGIS was more valuable in terms of the knowledge assessment than the skills and attitudes assessments. There was a statistically and educationally significant increase from the pretest to posttest on the knowledge assessment, but the overall posttest performance was below average. Overall performance on the skills assessment was also below average. Participants reported positive dispositions toward LOGIS and toward graphing, but no significant difference was found between the pre-instruction survey and the post-instruction survey. The value of LOGIS must be considered within the context that this study was the first iteration in the refinement of the LOGIS instructional application.

CHAPTER ONE

INTRODUCTION

Chapter Map

This chapter introduces the current study, and provides a framework and rationale for conducting the study. The following map describes the organization of the chapter:

- Introduction
    - Chapter map
    - Motivation
        - Practical motives
        - Research motives
    - Focus of the study
    - Significance of the study
    - Acronyms and definitions of terms
    - Summary

Motivation

Practical Motives

An online course at a major southeastern university uses Alberto and Troutman (2006) as its primary course textbook. Chapter four in the Alberto and Troutman (2006) text is titled “Graphing Data”, and in this chapter students learn how to create and interpret simple and cumulative graphs. Multiple baseline graphs are covered in Chapter five titled “Single-Subject Designs”. The pedagogy of both chapters is based on describing objects or features; stating facts or rules; and providing examples, summaries, and exercises. Because the graphing theme is distributed across chapters and the text does not intrinsically provide instruction, the professor of record created interactive instructional tutorials to initially augment but eventually replace the graphing chapters in the book. These tutorials were experimentally tested with students in prior semesters and although the posttest results were better than the traditional method of instruction, the tutorials were not as effective as had been envisioned.

One possible explanation for the modest performance of students on the tutorial posttest is the fact that the tutorials themselves did not require students to graph. The textual prompts and pictorial examples were evidently not powerful enough to cause individual students to produce an acceptable graph from the data provided. Although it seems obvious that graphing should be required with graphing instruction, such practice is often deemed uneconomical or impractical. Students are frequently expected to convert visual or auditory stimulus into new behaviors – learning. This is based on the common fundamental assumption in education that reading and/or lectures are sufficient for learning to occur.

The purpose of the current study was to create effective instruction for the graphing component of the specified course. This instructional application was named LOGIS, a recursive acronym that represents LOGIS Online Graphing Instructional Simulator. The general aim was for students to complete the instruction in LOGIS, pass the subsequent course quiz, and eventually pass the course. This study contended that the goal of creating effective instruction for this task was best realized if the instruction was paired with non-optional guided contingent practice, where forward progress within the practice task was dependent upon correct responses from the learner.

Research Motives

The creation of new instruction provided the opportunity to investigate the development process and engage in Developmental Research. The decision to create model-based instruction introduced the possibility of creating effective instruction and simultaneously analyzing the creation process. Using the Analysis Design Development Implementation Evaluation (ADDIE) model and detailing each step provided a foundation to comment on the effects of using a model-based approach to development, thus adding to the current literature.

There are many Instructional Systems Design (ISD) models and some, for example, the Dick and Carey model (Dick, Carey, & Carey, 2005), might have been more suitable for this particular task. The object of the study was not to compare models or develop another set of “best practice” guidelines, but to analyze the development process using the ADDIE model. The ADDIE model was chosen because it was the most generic and fundamental model, and comments on this model might extend to other derived models. The ADDIE model has five phases: Analysis, Design, Development, Implementation, and Evaluation. These phases guide the development of instructional applications by providing a framework for transparency and accountability within the development process. This study will comment on how well ADDIE achieved its goals.

Focus of the Study

The primary focus of this study was the model-based development of instructional software for a specific unit of instruction and the documentation and analysis of that process. Using a model provided the opportunity to build effective instruction and document the process thus increasing the transparency and repeatability of the study. This provided a basis for analysis and comment on the development process.

In addition to the practical development and documentation processes, this report will comment on the principles of response contingent progression in both instruction and guided practice.

Research Questions

This study focused on two non-experimental research questions:

1. How does the use of the ADDIE model influence the creation of the LOGIS instructional application?
2. Is LOGIS an effective form of instruction?
    a. Is LOGIS an effective form of instruction as measured by educationally significant differences in learners' performance from the pretest to the posttest on the Knowledge measure?
    b. Is LOGIS an effective form of instruction as measured by educationally significant differences in learners' performance from the pretest to the posttest on the Skills measure?

c. Is LOGIS an effective form of instruction as measured by educationally significant differences in learners’ attitude towards graphing from the pre to the post Graphing Attitude Survey?

Significance of the Study

This study is significant because firstly it answered the call for Developmental Research (Reeves, 2000a; Richey, Klein, & Nelson, 2004; van den Akker, 1999). This study engaged both the development process and the investigation of the development process, increasing the study’s practical and research scholarship value (Richey et al., 2004).

Secondly, this study implemented guided contingent practice and adaptive instruction principles, and both are aimed at increasing the effectiveness of the instruction. The literature (Kirschner, Sweller, & Clark, 2006) acknowledged the importance of guided practice and this report will provide scholarly comment on the issue.

Thirdly, the effectiveness of simulations is still questionable (de Jong & van Joolingen, 1998). This study cannot settle the debate, but this report will add relevant findings regarding the effectiveness of simulations. The instructional application that was developed contained a simulation component, and the analysis of this application yielded valuable insight into the effectiveness of simulators in a learning environment.

Finally, this study documented the development and evaluation processes, and used that documentation as a framework for comment on the use of the ADDIE model.

The comments are not simply advantages and disadvantages, but a narrative on how the use of the model affected the development process.

Acronyms and Definition of Terms

The acronyms in Table 1 are used within the current document. They increase readability without compromising understandability.

Table 1

Acronyms Used

Acronym	Meaning
PI	Programmed Instruction (PI) is a method of instruction emphasizing the linear and logical arrangement of reinforcement contingencies that cumulatively establish target behaviors.
ISD	Instructional Systems Design (ISD) is “the process for creating instructional systems” (Gagne, Wager, Golas, & Keller, 2005, p. 18).
ADDIE	Analysis Design Development Implementation Evaluation (ADDIE) is “The most basic model of the ISD process” (Gagne et al., 2005, p. 21).
LOGIS	Logis Online Graphing Instructional Simulator (LOGIS) is a recursive acronym describing the instructional application developed for the current study.

The terms in Table 2 are used within the current document. The definition for each term is based on the reviewed literature.

Table 2

Terms Used

Term	Meaning
Learning	"a process that results in a relatively consistent change in behavior or behavior potential and is based in experience" (Tomporowski, 2003, p. 7).
Developmental Research	"a situation in which someone is performing instructional design, development, or evaluation activities and studying the process at the same time" (Richey et al., 2004, p. 1099).
Simulation	"a model of some phenomenon or activity that users learn about through interaction with the simulation" (Alessi & Trollip, 2001, p. 213)
Attitude	"an internal state that affects an individual's choice of personal action toward some object, person, or event" (Gagne et al., 2005, p. 95).
Guided Contingent Practice	Repetition of logically arranged steps where progress is contingent upon correct incremental responses.

Summary

This chapter introduced the current study providing rationale, motive and justification. This Developmental Research study did not entail model creation or validation, but instead it described the process and effect of using a model to develop instructional software.

CHAPTER TWO

REVIEW OF THE LITERATURE

Chapter Map

This chapter is a review of literature that is relevant to both the theoretical and practical foundations of LOGIS. It describes the nature of Developmental Research and the value of this type of research. This chapter also establishes the precedence for the instructional techniques implemented in LOGIS and provides justifications for the inclusion of specific features into LOGIS. The following map describes the organization of the chapter:

- Review of the literature
    - Chapter map
    - Developmental Research
        - Historical perspective
        - The nature of Developmental Research
        - Issues and challenges in Developmental Research
        - Why conduct Developmental Research?
    - Learning
        - What is learning?
        - Experience and learning

    - Attitudes and learning
    - Learning outcomes
    - Relating learning perspectives to instruction
    - The assessment of learning
  - Instruction
    - Programmed instruction
    - Adaptive instruction
    - Intelligent tutoring systems
    - Simulations
    - The design of instruction
    - Addie
  - Graphing
  - Summary

Developmental Research

The current educational research literature recognizes Developmental Research as a unique and emerging area. Despite the acknowledged importance and relevance of this type of research, no clear consensus has been reached regarding the definition, scope, and overall character of Developmental Research.

This study includes a discussion of Developmental Research in an effort establish a base rationale and framework within which the study can be framed. A clear understanding of this type of research will focus the study, provide a basis for scholarly comment on relevant issues, and provide conceptual guidance in an area van den Akker (1999, p. 3) called "…rather confusing."

Historical Perspective

It is necessary to understand and appreciate the evolution of Developmental Research before attempting any meaningful dialog about its current and future status. Not only does background information provide the context for the current literature, it also provides a guide for future discourse.

To understand Developmental Research is to understand educational technology as a discipline. The recent calls for more and better Developmental Research scholarship from prominent figures (Reeves, 2000a; Richey et al., 2004; van den Akker, 1999) are understandable and expected given the history of educational technology. The almost cyclical nature of especially educational history (Cuban, 1986) presents an interesting dilemma where the case can be made that the call for Developmental Research is a wholly expected and necessary artifact of modern society and scholarship. Developmental Research, it can be argued, is the next link in the evolution of educational technology.

The intellectual framework for educational technology was developed in the early 1920s, but it was not until the 1950s that the academic programs and support structures were created (De Vaney & Butler, 1996). Educational technology emerged from the militaristic training model that emphasized both temporal and quantitative efficiency. Given the political climate, the emphasis on quantifiable science, and a positivist doctrine, it is understandable that education and consequently educational technology would have a deterministic bias. The scholarship of that time was very reflective of the social norms of that time. This is only one of many examples where scholarship parallels society.

Technological progress is also a lens through which the evolution of educational technology can be viewed. The growth of the discipline can be clearly seen as it evolves from early investigations of audiovisuals to more detailed research on current forms of technology. This reality has proven to be both positive and negative. While it is true that there has been some progress in the definition of the field and its relevance in academia, the current conclusion is that the field needs less media comparisons studies (Reeves, 2003). The debate between R. E. Clark (1983) and Kozma (1991) on the effect of media is both humbling and promising in that it suggested that decades of research are anecdotal at best. The mere presence and framing of this type of argument highlights how much the scholarship in educational technology has matured. On the surface, the current lack of structure and focus in educational technology seems to undermine the validity of the field, but as De Vaney and Butler (1996) pointed out, this might actually be to the credit of the field. They proposed the following

> The fact that past and present educational technology scholars have failed in this monolithic effort is to the credit of the field. Heterogeneous texts produced during the period under consideration and later provide a rich account of objects of study, theories engaged, methods employed, and audiences included. The written and oral texts considered here disclose a set of common goals but are diverse projects whose structures are contingent on historically accepted concepts and values. They reflect prevailing notions of learning theory and pedagogy, research methods, economic, military, and political values, and other elements of the social

> milieu in which they were produced. The iterations of names, concepts, assumptions, and theories in these texts not only promoted ideas but actually created truisms in the field for the time in which they were written. The value of these texts cannot be measured by sophisticated standards of current research, nor by highly evolved notions of learning theory, but by how they achieved their common goals when they were written. From whatever perspective these authors spoke, we might ask how well they made their objects of study intelligible to specific audiences at specific moments in time. The rhetoric with which they spoke and the discourses that spoke through them energized an audience of scholars, educators, and students to participate in a new field, educational technology. By any measure they were successful. (p. 3)

The Nature of Developmental Research

The nature of Developmental Research is tied to its definition. Seels and Richey (1994) defined Developmental Research as “the systematic study of designing, developing and evaluating instructional programs, processes and products that must meet the criteria of internal consistency and effectiveness” (p. 127). van den Akker (1999) proposed that Developmental Research had differing definitions that are sub-domain specific. Several sub-domains were examined to highlight what van den Akker termed “conceptual confusion” (p. 3). Within the Curriculum sub-domain, the purpose of Developmental Research is described as “…to inform the decision making process during the development of a product/program in order to improve the product/program being developed and the developers' capabilities to create things of this kind in future situations” (van den Akker, 1999, p. 3). Similarly, van den Akker quoted Richey and

Nelson’s (1996) aim of Developmental Research within the Media & Technology sub-domain, citing “improving the processes of instructional design, development, and evaluation … based on either situation-specific problem-solving or generalized inquiry procedures” (Richey & Nelson, as cited in, van den Akker, 1999, p. 4). The examples reveal different dimensions that all fall under the general heading of Developmental Research.

Developmental Research has had several labels over the years. It has been linked to Action Research and Formative Research/Evaluation to name a couple. Although the terms are often used interchangeably, Reeves (2000a) made a clear distinction between research with development goals and those with action goals.

Development Research, which Reeves (2000a) also referred to as Formative Research, is focused on “…developing creative approaches to solving human teaching, learning, and performance problems while at the same time constructing a body of design principles that can guide future development efforts” (p. 7). He emphasized the idea that development research addresses both the practical and theoretical issues involved in the learning process.

According to Reeves (2000a) Action Research is focused on “…a particular program, product, or method, usually in an applied setting, for the purpose of describing it, improving it, or estimating its effectiveness and worth” (p. 7). This type of research, Reeves suggested, is purported to solving a specific problem under select conditions within a limit time frame. Action Research, Reeves implied, does not have the same emphasis on theory as that of development focused research, thus it is not widely regarded as legitimate research. Reeves noted that under certain conditions, for example, reporting useful consumable results, Action Research could in fact be considered legitimate research.

Reigeluth and Frick (1999) discussed Formative Research and presented the argument

> …if you create an accurate application of an instructional-design theory (or model), then any weaknesses that are found in the application may reflect weaknesses in the theory, and any improvements identified for the application may reflect ways to improve the theory, at least for some subset of the situations for which the theory was intended. (p. 4)

They suggested that Formative Research is a subset of Developmental Research where the focus is on the development and testing of theories or models.

Action Research and Formative Research are only two of the terms associated with Developmental Research. Although they are perceived differently depending on the author, the common thread is the development and validation of useful instructional interventions.

Richey et al. (2004, p. 1099) presented a somewhat unifying theory of Developmental Research. They asserted that in its simplest form, Developmental Research can be either

- the study of the process and impact of specific instructional design and development efforts; or
- a situation in which someone is performing instructional design, development, or evaluation activities and studying the process at the same time; or
- the study of the instructional design, development, and evaluation process as a

whole or of particular process components.

There is a clear distinction between the development process and the studying of the development process. This distinction is very significant because it provides a framework for the characterization of Developmental Research.

Richey et al. (2004) proposed two categories for Developmental Research and they are differentiated by the nature of their conclusions. The first category (Type 1 research) includes studies that generally have context-specific conclusions. These studies involve a specific product or program design, and the development of that product or program. Typically, the design, development, and evaluation process of the entire instructional intervention is documented. Type 1 studies usually result in consumables, for example the impact of a program, or suggestions for improving a specific product. Richey et al. (2004) cited as an example “McKenney’s (2002) documentation of the use of CASCADE-SEA, a computer-based support tool for curriculum development” (p. 1102). Although McKenney and van der Akker (2005) confirmed that the study had a developmental approach, they also offered the following caveat “The research approach in this study may be more specifically labeled as formative research, since it involved the actual design and formative evaluation of a program” (p. 47). While the distinction appears to be trivial, it would be interesting to know if the definition affected the study to the degree that the distinction was worth mentioning.

The second category (Type 2 research) includes studies that typically have generalized conclusions. Although they may involve the analysis of the design and development of a program or product, Type 2 studies generally occur after the development process is complete. Type 2 research studies are purposed at producing knowledge, thus it is not uncommon for these studies to focus on model creation and validation. These studies usually produce, for example, new or enhanced models, or evidence of a model's validity. Richey et al. (2004) cited as an example of Type 2 research, a study conducted by Jones and Richey (2000). The study was based on the principle of Rapid Prototyping, and they proposed a revised ISD model that included Rapid Prototyping. Although the study was characterized as Type 2, Jones and Richey (2000) noted "Many view RP methods essentially as a type of formative evaluation that can effectively be used early and repeatedly throughout a project (Tessmer, 1994)" (p. 63). Once again, the importance of the definition was acknowledged.

Many different research procedures can be used when conducting Developmental Research. The examination of a process, as is the case in Developmental Research, affords the possibility of gathering rich data which in turn increases the validity of the study. Considering that the setting is usually real-world based, these studies often employ traditional quantitative procedures and additionally, they may also include qualitative aspects. Given the goals of Developmental Research, it becomes clear that the task of describing processes requires traditional and alternative perspectives. Many different instruments can be used to collect data, and many techniques can be used to analyze and report the data. A very good example of this is the CASCADE-SEA project (McKenney & van der Akker, 2005), where 108 instruments were used to collect data and several different procedures were used to analyze and report the data.

Issues and Challenges in Developmental Research

Conducting a study that has developmental goals is not a trivial task. While the potential exists for rich data and significant conclusions, rich data collection and analysis take time. Researchers acknowledge that this type of research requires an appreciable investment in time, and often forgo Developmental Research studies instead focusing on scholarship that is quicker to complete and more importantly quicker to publish (Reeves, 2000a). In an environment where researchers must publish or perish, Developmental Research is often avoided.

Developmental Research like educational research as a whole seems to be almost disjoint from its stakeholders, namely educators. The perception that educational research is only useful to scholars is not totally without merit. D. W. Miller (1999) noted “Some scholars contend that education research can boast plenty of solid, useful findings about learning and reform. The problem is not that the research isn't good, they say, but that it doesn't find its way into the classroom” (p. A18). In that scathing article, D. W. Miller suggested that the education discipline is not only failing to shape classroom realities, but its research “is especially lacking in rigor and a practical focus on achievement” (p. A17). This position is supported by Reeves (1995) when he characterized significant portions of educational research as “pseudoscience” (p. 6).

Although the failings of education research are evident, most scholars acknowledge that hope exists. Recently, prominent scholars (Reeves, 2000a; Richey et al., 2004; van den Akker, 1999) have called for more Developmental Research to be conducted. They agreed that Developmental Research is one avenue through which academic and practical solutions can be found. Obviously Developmental Research will not solve all the problems in educational research and subsequently education, but considering what is at stake, Reeves (2000a) put it best when he said “Given the poor history of other approaches, I am increasingly convinced that if instructional technologists want to contribute to meaningful educational reform, they should pursue development goals" (p. 11).

Why Conduct Developmental Research?

There are many reasons why researchers should conduct Developmental Research. Consider that more and better research with development goals will essentially increase the credibility of the field. Also, consider that this type of research is more apt to bridge the gap between the scholar and the practitioner thus increasing the value of research. Although those reasons are very good in and of themselves, perhaps the best reason to conduct Developmental Research is because it is socially responsible. Most of the literature reviewed for the current document share the common theme that Developmental Research is simply the right thing to do. The sentiment is best expressed by Reeves (1995) when he suggested "It would seem that we stand a better chance of having a positive influence on educational practice if we engage in Developmental Research situated in schools with real problems" (p. 465), and concluded that "We cannot afford to lose another generation of researchers to the pursuit of research for its own sake. We must be more socially responsible" (Reeves, 2000b, p. 27).

Learning

This section examines some of the discourse pertaining to learning. It is important to address the issue of learning on both a concrete and an abstract level because the perceptions of what learning is dictate the implementations of instruction and the procedures for assessment. If the terminal objective is the production of an instructional intervention, then a clear understanding of learning is a logical starting point.

This discussion will form a part of the framework for this study. A clear definition of learning will be generated, thus facilitating the creation and development of the instructional intervention.

What is Learning?

In general, people agree that learning is important, but the causes, processes, and consequences are still contentious issues. The definition of learning changes depending on perspective, but there are common threads and similar themes across different perspectives. This review will not include a detailed discussion of different learning perspectives but a general synopsis of five of the most common orientations to learning can be found in Merriam and Caffarella (1999, p. 264).

The behavioral and the cognitive perspectives will be the primary focus for this review. The reason these two were selected is that they represent fundamentally different but similar propositions; the behavioral emphasis on the overt environment juxtaposed to the cognitive emphasis on covert mental processes. Although different, a thorough examination would reveal that behavioral and cognitive positions contain many common threads. McDonald, Yanchar, and Osguthorpe (2005) suggested that

> In most fundamental respects, however, cognitivism and behaviorism are virtually indistinguishable—they are both rooted in a deterministic (mechanistic) ontology that views human action and learning as the necessary output of environmental inputs and biological conditions; and both are based on an empiricist epistemology that views the mind—including behavioral repertoires, schemas, mental models, and so on—as gradually constructed over time through the mechanistic processing of sensory impressions. (p. 91)

Scholars like Piaget, influenced by both behavioral and cognitive schools concluded that learning is affected by both internal and external agents (Merriam & Caffarella, 1999).

The early behavioral definitions of learning focused on learning as a product. Jarvis (1987) provided an example when the he quoted Hilgard and Atkinson's definition of learning as "a relatively permanent change in behaviour that occurs as a result of practice" (p. 2). For the purpose of scientific inquiry this definition is clear and concise, but the simplicity of the definition led to criticisms that questioned, for example, whether or not the behavior had to be performed in order for learning to have occurred, and also whether or not potential for change was taken into consideration (M. K. Smith, 1999). Jarvis, a critic of the behavioral definition, proposed an expansion to the behavioral definition and suggested that learning is "the transformation of experience into knowledge, skills and attributes, and to recognise that this occurs through a variety of processes" (p. 8).

Most of the critiques of the behavioral position are structured around the pervasiveness of the mind, that is, to what extent the mind controls the individual. While the arguments are philosophical in nature and not extensively treated in this review, it is important to note that most critiques of the behavioral position are erringly structured on trying to fit behaviorism within the cognitive scope. Burton, Moore, and Magliaro (2004) surmised that

> Skinner's work was criticized often for being too descriptive—for not offering explanation. Yet, it has been supplanted by a tradition that prides itself on qualitative, descriptive analysis. Do the structures and dualistic mentalisms add anything? We think not. (p. 27)

Emurian and Durham (2003) suggested that within the behavioral context, the antecedents and interactions were sufficient in explaining learning. They classified their approach as "atheoretical" and asserted that

> It is *atheoretical* in that it will focus on the interactions themselves as the explanation of the antecedents to knowledge and skill acquisition, and it will rely on those antecedents, rather than external explanatory metaphors, such as cognitive models, to explain the process and outcome of learning. (p. 679)

Jarvis (1987), in dissecting the behavioral definition of learning, proposed several critical flaws, one of which was:

> First, if a person can be taught to think critically and also to be autonomous, then it is difficult to maintain that what is going on within a person in subsequent situations is merely the result of the environment, or determined by previous experiences. (p. 4)

Skinner (1972) proposed that the functions of autonomous man could be attributed to the controlling environment, and Jarvis (1987) considered this the point of contention. It should be clarified, however, that cognitive structures cannot be arbitrarily assigned to behavioral definitions. Skinner (1957) viewed thinking as

> The simplest and most satisfactory view is that thought is simply *behavior* – verbal or nonverbal, covert or overt. It is not some mysterious process responsible for behavior but the very behavior itself in all the complexity of its controlling relations, with respect to both man the behaver and the environment in which he lives. (p. 449)

The concept "critical thinking", which Jarvis mentioned, differs between the two perspectives. This fundamental difference is not accounted for within the critique as Jarvis used the cognitive values of thinking as a tool to comment on the behavioral position.

Cognitivists extended the behavioral definition of learning to include mental convert processes and the "capacity" (Schunk, 2000, p. 2) to learn. Currently, the generally accepted cognitive definition of learning emphasizes learning as a process. Tomporowski (2003) quoted Zimbardo and Gerrig's definition of learning as "a process that results in a relatively consistent change in behavior or behavior potential and is based in experience" (p. 7). This definition implies three characteristics. First, the term learning is only used in situations where an overt change in behaviors occurs consistently over a given time frame. Secondly, because learning is defined as a covert process, the behaviors must be demonstrated to prove that learning has occurred. Once proven, the behavior will become a relatively permanent part of the learner's repertoire. Finally, learning can only occur with practice or experience (Tomporowski, 2003). These conditions are direct target goals for the current study.

Although Burton et al. (2004) proposed a scientific definition of learning "…a function of building associations between the occasion upon which the behavior occurs (stimulus events), the behavior itself (response events) and the result (consequences)" (p. 9), this study will define learning as "a process that results in a relatively consistent change in behavior or behavior potential and is based in experience" (Tomporowski, 2003, p. 7). The latter definition captures both the behavioral background and current cognitive influences of learning, and provides an opportunity for the measurement of overt learning artifacts.

Experience and Learning

Traditionally, experiential learning has been used in two ways: learning via contact with the phenomena being studied, and learning via the events of life. The literature presents discussion on the social construction of learning, and reflective and non-reflective learning (Jarvis, Holford, & Griffin, 2003). Although the arguments are purposed at teasing out the attributes of learning thus generating a clear definition, Jarvis et al. (2003) admitted that "…all learning is experiential and so any definition of learning should recognize the significance of experience" (p. 67).

One of the central tenets of the current study is that experience is vital to learning, but the wider social context of life experiences is not considered. This study is primarily concerned with the physical connection between a learner and a target behavior; the relationship being experience. The Kolb and Fry Model (Jarvis et al., 2003, p. 59) is shown in Figure 1. It was developed in 1975 and highlights the importance of concrete experiences in the learning cycle.

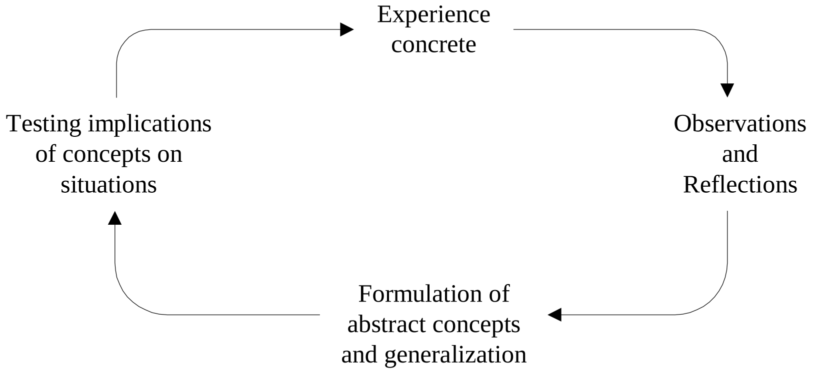


Figure 1. The Kolb and Fry learning model.

Although the model is critiqued as being too simplistic, it is generally considered the acceptable foundation for experiential learning discourse. The Kolb and Fry model is also the basis for other more complex models (Jarvis et al., 2003, p. 59) that seek to address the importance of secondary life experiences.

The importance of experience in learning is clear. What remains uncertain are the attributes of the experiences that optimally produce learning, and the conditions under which they occur. Ericsson, Krampe, and Tesch-Römer (1993) presented evidence that suggested expertise could be explained by extended deliberate practice as opposed to innate characteristics. While they admitted that the relationship is subject to various confounds, their study showed that expert behavior can be reliably linked to extended deliberate practicing. Even more important was the assertion that immediate informative feedback is necessary for accurate performance, and the lack of feedback seriously inhibits learning even for highly motivated learners. Pimentel (1999) went further and noted that learning environments must have high levels of interaction. Interestingly, Pimentel stated that “the environment does not provide a lesson in an explicit fashion, rather it is something that humans do naturally. The environment simply provides the means for lessons to be learned” (p. 77), hinting at the usefulness of unguided non-explicit instruction.

Pimentel (1999) developed a complex virtual environment (LeProf) that provided learners with experiences that were both meaningful and transferable. The simulation allowed learners to manually input circuit parameters and experience different outputs. The interactive nature of the simulation was reported as successful because learners expressed positive attitudes towards the simulation. Interactive simulations like LeProf have benefits, but the current study contends that explicit instruction must be paired with a simulation for meaningful learning to occur. It is not sufficient to present an interactive environment; instruction and guidance are necessary components of learning. Kirschner et al. (2006) supported the need for direct instruction guidance stating

> After a half century of advocacy associated with instruction using minimal guidance, it appears that there is no body of research supporting the technique. In so far as there is any evidence from controlled studies, it almost uniformly supports direct, strong instructional guidance rather than constructivist-based minimal guidance during the instruction of novice to intermediate learners. Even for students with considerable prior knowledge, strong guidance while learning is most often found to be equally effective as unguided approaches. Not only is unguided instruction normally less effective, there is evidence that it may have negative results when students acquire misconceptions or incomplete and/or disorganized knowledge. (p. 83)

It becomes evident that highly interactive simulation environments must contain elements of direct guided instruction. While experiential learning advocates placing the learner in contact with the phenomena to be studied, it is not sufficient to simply facilitate the contact. The learner must be guided within the medium such that important aspects are highlighted and irrelevant artifacts ignored. R. E. Clark (2005) addressed this issue, developing the Guided Experiential Learning (GEL) process for completing and validating the design and development aspects of instruction.

The current study fundamentally contends that experience in the form of concrete contact is vital in learning but this experience must not only be opportunistic, it must be required. It is not sufficient to simply present the opportunity to practice or engage experiences; the practice must be required and contingently guided.

Attitudes and Learning

The relationship between learning and attitude is very important. Jarvis (1987) suggested that learning is "the transformation of experience into knowledge, skills and attributes" (p. 8). Knowledge and skill are measurable attributes, and are a genuine fit within the behavioral framework. Attitude, in its native form does foster accurate assessment because it is essentially a metaphor describing the state of, in this case, a learner. A definition consistent with the theme of this study must be developed if the question of attitude is to be addressed.

The study of attitudes is a major focus in the behavioral and psychological sciences. The volume of knowledge and research in this area is beyond the scope of this review but at minimum the scholarship will provide a foundation for defining and describing attitudes.

Generally, attitudes are described in affective terms, for example, like/dislike and good/bad. An attitude towards an object is determined by subjective values of an object's attributes and strength of the associations (Ajzen, 2001). An individual may have many different beliefs towards a single object, but only those beliefs that are readily available in memory influence attitudes at a given moment (Ajzen, 2001), thus the temporal nature of attitudes is exposed.

In describing attitude formation, Crano and Prislin (2006) stated

> Today, most accept the view that an attitude represents an evaluative integration of cognitions and affects experienced in relation to an object. Attitudes are the evaluative judgments that integrate and summarize these cognitive/affective reactions. These evaluative abstractions vary in strength, which in turn has

> implications for persistence, resistance, and attitude-behavior consistency. (p. 347)

This definition highlights the inherent subjectiveness of attitudes and further clarifies the notion that an attitude is a collection of judgments affecting behavior towards an object. Crano and Prislin (2006) further discriminated between attitude formation and attitude change by describing the Dual-Process model of attitude change

> Dual-process models hold that if receivers are able and properly motivated, they will elaborate, or systematically analyze, persuasive messages. If the message is well reasoned, data based, and logical (i.e., strong), it will persuade; if it is not, it will fail. Auxiliary features of the context will have little influence on these outcomes. However, if message targets are unmotivated (or unable) to process a message, they will use auxiliary features, called “peripheral cues” (e.g., an attractive source), or heuristics (e.g., “Dad’s usually right”) to short-circuit the more effortful elaboration process in forming an attitudinal response. Such attitudes are less resistant to counterpressures, less stable, and less likely to impel behavior than are those formed as a result of thorough processing. (p. 348)

This description introduces both the individual ability differences in learners and the motivational factors involved in attitude change. It is clear from the description that a capable and willing learner will change an attitude if the message is sufficiently strong.

The development or change of an attitude might occur over a period of time, or after isolated contact. If a new attitude is established and it is strong, it will be stable over time, it will be persuasion resistant, and most importantly it will be a predictor of future behavior (Ajzen, 2001).

It is important to remember that the exact relationship between behavior and attitude is still unknown, but it is widely accepted that each influences the other (Ajzen, 2001). The many variables involved in attitude formation or change makes the measurement of this attribute very difficult.

Gagne et al. (2005) defined attitude as “an internal state that affects an individual’s *choice of personal action* toward some object, person, or event” (p. 95). The importance of this definition is that it introduces the measurable construct *choice*. While this definition admittedly does not capture the entire scope of attitudes (Gagne et al., 2005), it is a consistent subset of the current literature and it is directly applicable to the current study.

The Gagne et al. (2005) definition can be viewed as a subset of the definition proposed by Crano and Prislin (2006) under two conditions. Firstly, “evaluative judgments” and “cognitive/affective reactions” are internal constructs and can be correctly labeled internal states that are removed from casual analysis. Secondly, the concepts of evaluate, integrate, and summarize are all behaviors directed towards a target recipient. Clearly, attitudes are a combination of complex behaviors and the result of many variables. The Gagne et al. definition, although less precise than the definition by Crano and Prislin, captures these complex behaviors and their internal antecedents and consequences in a measurable way – choice. The aim was not to trivialize or minimize the contributions of internal agents, but to develop a context through which attitudes can be objectively assessed while reserving cognitive and affective comments for the authors who work within those fields (Gagne et al., 2005, p. 94). The current study will use the

Gagne et al. (2005) definition of attitude as a base to describe the measurement and subsequent analysis of attitudes.

It is tempting to classify attitudes as predictors of future behavior, but that is only partially correct (Ajzen, 2001). Only a strong attitude will predict future behavior with acceptable accuracy. Weak attitudes are subject to external confounds and thus they are not an accurate measure of future behavior. Within this context, the current study does not measure attitude as a predictor of future behavior, rather, given the temporal nature of attitudes, it measured attitude at a single point in time. That measurement will not be used as a predictor of behavior, but rather as a point of reference to describe the possible development of future behavior.

Learning Outcomes

Learning outcomes were classified based on Gagne et al.'s (2005) types of learning outcome. This classification scheme was chosen because it is consistent with the general theme of the study and it logically fits with the other parts of the current study that are based on Gagne's work. Figure 2 shows Gagne et al.'s (2005) categorization of types of learning outcomes.

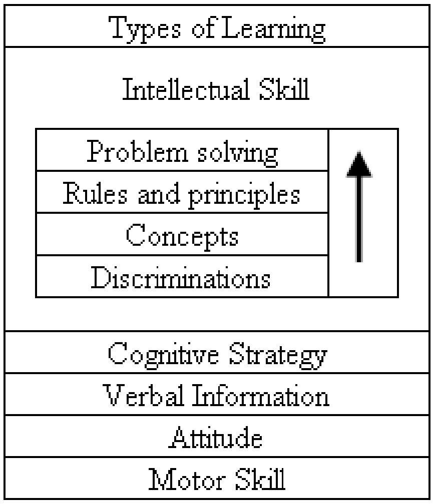


Figure 2. Gagne's categorization of learning outcomes.

A description of each outcome is provided followed by an example performance indicator.

Intellectual Skill

A class of learned capabilities that allows an individual to respond to and to describe the environment using symbols, for example, language or numbers. This class is divided into hierarchical levels where each level is a prerequisite for the next.

Discriminations. Discrimination refers to the ability to identify differences in stimuli based on a given dimension. The learner must be able to discriminate between, for example, the ordinate and the abscissa, indicating that the learner can distinguish similar and different attributes of a stimulus.

Concepts. A concept allows the learner to classify stimuli based on general or common attributes. When a learner identifies the properties of an object that make it a member of a class, it is an indication that the learner has acquired the concept governing that object.

Rules and principles. Rules or principles are statements that describe the relationship among concepts. Most complex behaviors, for example swimming, are based on rules, thus engaging in complex behaviors is an indication of rule acquisition. It is not sufficient to state rules, they must be applied.

Problem solving. Problem solving is a process leading to an instructional outcome. It is not the outcome itself. Rules are sometimes used in problem solving, but this is not mandatory. Discovery Learning is an example of problem solving. Most problem solving involves the use of complex rules formed from simpler rules. Taken together, they can be used to solve a specific problem.

Cognitive Strategy

A cognitive strategy is an internal process where learners engage the way they remember, learn, attend, and think. There are many types of cognitive strategies including rehearsal, elaboration, and organizing, but they are all methods that facilitate self-modification. Cognitive strategies are internal processes and cannot be readily observed. They must be inferred by querying other intellectual skills or obtained via self-reports.

Verbal Information

Verbal information or declarative knowledge provides a foundation for learners to build other skills. Verbal information knowledge is built on information which is in turn built on data. For example, the time is 9:30am (data) and behavior occurs every hour on the hour (information) leads to the declarative knowledge that the behavior is not occurring at this time.

Attitude

Attitude, defined in the "Attitudes and Learning" section of this document, is an internal state that affects an individual's choice of personal action toward some object, person, or event. An attitude can be measured by observing the choices learners make under certain conditions.

Motor Skill

Motor Skills are learned capabilities reflected in bodily movements. Practice is a key issue in developing motor skills, and performance of the skills under specified conditions indicates acquisition of that skill.

Gagne et al. (2005) pointed out that although a majority of instruction includes most or all of the categories, classifying learning can be useful because grouping objectives reduces design work and facilitates optimal instruction sequencing and planning.

Relating Learning Perspectives to Instruction

Hartley (1998) proposed a set of principles that guide learning and consequently instruction. A detailed discussion on instruction is included in this review, but Hartley listed principles that bridge theory and practice. These principles provide a framework where the abstractions of the theoretical perspectives can be transformed into concrete usable artifacts. Although this is not an exhaustive set of principles, it does provide a first step in determining what instruction should look like based on behavioral and cognitive perspectives. Key behavioral principles emphasized during the learning process include:

- Activity is important. The probability of learning increases when learners are actively involved in the learning process.

- Repetition, generalization, and stimulus discrimination are important if learning is to incorporate transfer to new contexts.
- Reinforcement, especially positive reinforcement, is a powerful motivator.
- The presence of objectives aids the learning process.

Key cognitive principles emphasized during the learning process include:

- Instruction should be logically ordered and well structured.
- Well-organized instruction is easier to learn than poorly organized instruction.
- The way the instruction is presented is important. Perceptual features of the task are important thus it might be a good idea to, for example, give general outlines of tasks to be covered before the instruction begins.
- Prior knowledge is very important, and learning that fits within or extends the learner’s current knowledge base will probably be more positively received.
- Individual differences in, for example intellectual ability or personality, affect learning.
- Cognitive feedback regarding success or failure increases the probability of learning.

It is evident that there is considerable overlap between the principles, thus it is possible to incorporate many or all of the principles into one instruction strategy.

The Assessment of Learning

Jarvis et al. (2003) discussed assessment as an extension of learning perspectives. They outlined the importance of assessment as

> …how people’s learning is assessed determines to a large extent what learning they think is important. Assessment is not therefore a ‘neutral’ technique for measuring the performance of learners. It has become a central feature of how education and training are organized in almost every society. (p. 158)

Although they did not treat specific assessment techniques in detail, they provided key features that broadly reflect the development of assessment literature and are pertinent to the current study:

- Formal and informal assessment
    - Formal assessment is more purposeful and organized than informal assessments
- Formative and summative assessment
    - While formative assessments are used to determine or alter current teaching or learning, summative assessments reflect what has been learned at the end. It must be noted that in practice, assessments are usually conducted for both formative and summative reasons.
- Measurement
    - Assessment of learning may include some numeric representation of achievement, or in certain circumstances labels are more appropriate.
- Judgment
    - Often, judgments are made regarding the level of mastery that a learner has achieved. An example of this is the allocation of a letter grade, for example an A, as an indication that the learner has mastered a particular content. The teacher in this case makes a judgment as to mastery level of

    the learner.

- Validity
    - Assessments that are valid measure only what they are supposed to measure. If a learner is to be assessed in a particular area of a subject, then the assessment instrument must measure that particular content area. Valid assessments can be used to accurately determine what mastery level the learner has achieved in a particular area.
- Reliability
    - A reliable assessment will return consistent results for different learners who perform similarly. This means that across all learners, the instrument will return similar scores for learners at the same performance level.

Instruction

Programmed Instruction

Skinner

> Probably no single movement has impacted the field of instructional design and technology more than Programmed Instruction. It spawned widespread interest, research, and publication. It was then placed as a component within larger systems movement, and finally, it was largely forgotten. (Lockee, Moore, & Burton, 2004, p. 545)

The term Programmed Instruction (PI) was probably the result of Skinner’s 1954 paper entitled “The Science of Learning and the Art of Teaching”. Skinner (1954) was mostly a concerned reaction to his daughter’s classroom realities at the time, but it set the stage for Skinner’s comment on the science and technology of human behavior.

Skinner (1958) formally proposed the programming of instruction as a way of increasing learning. He noted that education needed to become more efficient to deal with the steady population growth. Audio-visuals aids were being used to supplement instruction, but Skinner felt that although content could be delivered via visual aids, “There is another function to which they contribute little or nothing. It is best seen in the productive interchange between teacher and student” (Skinner, 1958, p. 969). Skinner believed that instruction could be automatically and mechanically delivered while maintaining the teacher/learner interchange in a tutorial style environment. His aim was to create an environment where the learner was not “a mere passive receiver of instruction” (p. 969).

The Sidney Pressey Teaching Machines of the 1920s, which Skinner used as a foundation, had several features that Skinner believed to be paramount. The most important feature of the teaching machines was that they permitted learners to work at their own pace, and facilitated learning by providing immediate feedback to the learner. Although the Pressey machines failed in part to what Skinner called “cultural inertia” (Skinner, 1958, p. 969), the principles of immediate feedback and the learner as an active participant in instruction remained.

Skinner’s idea of teaching machines was less an instrument and more of a set of principles to bring learning under the control of specific stimuli. His proposed machine had several important features that reflected Skinner’s view on learning. The machine should require that the learner compose as opposed to select responses. The aim according to Skinner is to promote recall in lieu of recognition. Another feature Skinner mentioned was the sequencing of learning contingencies such that a learner would traverse a set of small steps that would lead to the desired terminal behavior. It is within this context that Skinner introduced the "frame" as a presentation of visual material that required a response, and that response would then elicit immediate appropriate feedback. A frame, discussed later in this section, would contain content that would be differentially reinforced, bringing verbal and nonverbal behaviors under the control of specific stimuli.

The teaching machine was not designed to teach, but rather "It simply brings the student into contact with the person who composed the material it presents" (Skinner, 1958, p. 973). The machine Skinner envisioned would facilitate mass instruction while retaining the "good tutor" quality that Skinner insisted was important. Skinner (1986) envisioned that the personal computer could, for the first time, truly facilitate mass instructional while retaining the individualized characteristics of a personal tutor.

Instructional Characteristics

PI can be specified as a "sequential arrangement of reinforcement contingencies that cumulatively establish terminal repertories – as well as their stimulus control" (Davis, Bostow, & Heimisson, in press). PI encompasses several principles and techniques but no general consensus exists as to a standard approach to PI, hence the reference to PI as an art form by Skinner (Skinner, 1958). Lockee et al. (2004) described some of the commonalities that exists across differing approaches to PI. They mentioned the following components:

- Specification of content and objectives
    - Determining the content to be taught, including terminal behavior and measurable objectives.
- Learner analysis
    - Gathering data, for example demographic data, on the learner in an effort to customize the instruction.
- Behavior analysis
    - Analyzing the relevant behaviors to facilitate the sequencing of the instruction.
- Selection of a programming paradigm
    - Determining the sequencing technique to be used. Examples of techniques include linear, branching, and intrinsic programming.
- Sequencing of content
    - Several sequencing techniques exist, including a general linear progression based on objectives, and the RULEG system developed by Evans, Glaser, and Homme in 1960.
- Evaluation and revision
    - Evaluating learner responses at the end of the instruction in an effort to fine-tune the content and sequencing.

Completion of the components increases the probability of a successful instructional program, but each component is not required.

Several concepts are important when constructing PI tutorials. At the fundamental level, the programmer is concerned with the creation of individual “frames”, and the sequencing of those frames. A frame is a singly displayed presentation of visual material (Skinner, 1958), requiring an overt response. The sequencing of the frames affects the effectiveness of the overall tutorial. The aim of the tutorial is to “shape” a desired behavior and to accomplish this goal the tutorial must differentially reinforce the correct forms of the desired behavior. In essence, PI moves a behavior from a low probability of occurrence, to a high probability of occurrence via the shaping process.

A tutorial can contain several different techniques that help in the shaping process. Techniques like fading, the use of thematic or formal prompts, the use of copy frames, and priming all help bring the learner under the control of specific stimuli. These common techniques are described below:

Prompts. The use of prompts is central in PI. Prompts act as a sort of hint for the learner increasing the probability that the desired behavior will be emitted. Prompts can be formal or thematic. Formal prompts include a form of the desired behavior. For example, help letters for a missing key word. In this case, the formal prompts increase the probability of the desired response, that is, the construction of the key word.

Thematic prompts generate desired responses by cuing the learner via contextual clues. The use of synonyms, antonyms, and common comparisons are all strategies that take advantage of context to provide thematic prompts.

Fading. Fading involves the removing of clues from individual frames as the tutorial progresses. The withdrawal of formal or thematic prompting clues helps the learner to become independent of the need for such clues. Low probability responses become high and are emitted without the need for artificial prompting or clues.

Copy frames. Copy frames are unique in that they contain the desired response within the frame itself. Initially a tutorial may require a learner to emit a response or behavior that has not yet been established. The copy frame presents the response within its visual presentation and requires that the learner repeats or copies that response. Copy frames can be used to prime responses so that they can be shaped later in the tutorial.

Priming. Priming involves the presentation of a stimulus that the learner must independently produce at a later time. Priming a response early in a tutorial is necessary if that response is to be shaped and later produced with and without clues. It can be summarized and contextualized with the following paragraph:

Programmed Instruction is an effective teaching tool. Desired responses, which initially have a very low probability of occurrence, are primed using copy frames that require that the learner simply repeat the provided stimulus. After priming, the probability that the currently weak response will occur is increased by a process called shaping. Shaping involves differentially reinforcing correct forms of the response. The learner is successively reinforced for correct responses, and these responses become progressively difficult and prompts are systematically removed. The incremental withdrawal of prompts is called fading, and it is used to help transform a low probability response to a high probability response. Finally, a learner must construct the response without the aid of either formal or thematic prompts, thus the response is now under the control of stimulus presented in the tutorial.

Skinner (1958) admitted that programming instruction was an art form, and envisioned a day when it could be less an art, and more a technology. He did find conciliation in the fact that art form or not, “it is reassuring to know that there is a final authority – the student” (p. 974).

One criticism of PI is that this type of instruction is only effective for low achieving learners. The validity of this claim remains uncertain because the definition for “low achieving” is not usually defined within the context of the criticism. There are, however, situations where PI is effective and situations where it is not. Emurian and Durham (2003) proposed that

> Programmed instruction approaches may be best suited for students who have not mastered the art of studying, and one important benefit of completing a programmed instruction tutoring system is that it teaches a learner how to acquire knowledge independent of the domain. The ultimate objective, then, of a programmed instruction tutoring system is to free the learner for advanced study undertaken with traditional forms of knowledge codification, such as a book (p. 694)

Someone who has not learned how to study cannot be immediately labeled as a low achiever, but that person has a high probability of performing poorly because they have not learned to study. PI in this instance can be effective.

Learners who are motivated and have mastered self-management and studying can also benefit from PI. These learners are more likely to quickly and successfully advance through PI frames, thus learn at a fast rate. An interesting proposition is that for motivated learners, learning is independent of technological support. This would suggest that value of PI or another instructional system is determined by the nature of the learner (Emurian & Durham, 2003).

Research on PI

Analysis of research on PI and on educational technology as a whole is often preceded with disclaimers. Lockee et al. (2004) included a sub-section (20.4.1) titled “A Disclaimer” where they proceeded to characterize historical research with terms like “lacks validity” (p. 552) and “buyer beware” (p. 553). These characterizations represent the current view scholars have of educational technology. Today, most scholars agree that historical research on PI is littered with some combination of wrong assumptions, bad design, incorrect analysis, or just plain insignificance (De Vaney & Butler, 1996; Gredler, 2004; Reeves, 2000a). PI, as a consequence of the behavioral movement, is assumedly one of the main culprits because “Programmed Instruction (PI) was an integral factor in the evolution of the instructional design process, and serves as the foundation for the procedures in which IT professionals now engage for the development of effective learning environments” (Lockee et al., 2004, p. 545).

Historical PI research studies were often comparison studies, where PI was compared to traditional instructional methods. Many of these studies had confounding variables and frequently suffered from novelty effects and sampling errors. Lockee et al. (2004) cited several examples where obvious sampling or design errors were made, thus justifying the need for caution when analyzing the results. Most of the comparative studies found that learner performance after instruction using PI was either better or the same as traditional instruction, hence “no significant difference”. Most studies, however, noted that PI tended to produced equal or better results in less time. The no significant difference mantra that would eventual label the entire field, would in this case be positive because PI could produce at least the same results as traditional methods in less time and at a cheaper cost. This reality would become the primary reason why PI was adopted and adapted for use by the military, an area where efficiency and mass distribution were premium.

Early PI research contains documented errors and flaws, but the effort did yield several principles that are still relevant today. Lockee et al. (2004) described several key components of PI that are general and can be associated with any instructional intervention. They listed the key components as mode of presentation, overt versus covert responses, prompting, confirmation, sequence, size of step, error rate, program influence by age level, type of response - constructed vs. multiple choice, and individual versus group uses. Three of these components are worth further description because they have significant implications for current instructional practices.

Overt Versus Covert Responses

The issue of overt versus covert responding is a central issue in instruction. Is an observable response necessary for meaningful instruction? Lockee et al. (2004) cited various studies where researchers found no significant differences in overt versus covert responding and several studies where significant differences were found. They cited two studies that were done at Western Michigan University, a doctoral dissertation done by S. Shimamune in 1992, and a master's thesis by P. Vunovick in 1995. Those two studies found no significant differences in the effectiveness of instruction requiring overt responding compared to those that used covert responses. Lockee et al. also cited Tudor and Bostow (1991), and Tudor (1995) and these found differences where overt responding significantly outperformed covert responding. The significance of the above two pairs of studies is that they were all used as the basis for a study by M. L. Miller and

Malott (1997). M. L. Miller and Malott hypothesized that the performance-based incentives that were not a part of the Tudor studies could explain the difference between the previously mentioned two pairs of studies and others studies that were grouped respectively. M. L. Miller and Malott concluded that "the superiority of overt responding is a robust enough phenomenon to occur even when an incentive is provided for performance on the posttest" (p. 500). Although the M. L. Miller and Malott study validated the need for overt responses in instruction, closer inspection reveals that the study could be criticized for several reasons. Firstly, the sample was self-selected thus exposing the study to self-selection bias. Secondly, the final sample sizes were small. Finally, it appears that both groups received some form of incentive, but the exact distribution and criteria of the incentives were not clearly described in the journal article. These cautions do not invalidate the results, but certainly reduce the authority of the study.

Kritch and Bostow (1998) examined the issue of constructed responses in an effort to investigate the importance of a high rate of response construction. They found that high-density responding (overt) significantly out-performed low-density responding (covert), and the performance gains were also reflected in the applied task that was assigned. Kritch and Bostow observed that there was no statistical difference between high-ability and low-ability learners, where the ability measure was based on self-reported grade point averages. In addition, they found that higher learner gains occurred when instructional time was greatest, and carefully noted that the results were expected to be generalizable. None of the criticism of M. L. Miller and Malott (1997) can be applied to the Kritch and Bostow study.

The example sequence that was presented above highlights the fact that research on this topic remains inconclusive and the general recommendation from authors is that more work needs to be done in this area.

Constructed Response vs. Multiple-Choice

The nature of responses is an important consideration in the design of instruction. Several studies in this area found no significant difference between instruction requiring constructed responses and those using multiple-choice selections. The theoretical issue in this area is whether the distracters in multiple-choice questions have an adverse effect on shaping. Currently, most instructional interventions use a combination of the two, using each where applicable. Lockee et al. (2004) did not mention the Kritch and Bostow (1998) study in this section, but the latter does highlighted the benefits of constructed responses. Although they did not make comparisons with multiple-choice, Kritch and Bostow asserted that "…frequent overt constructed responding within instructional contingencies is a critical design feature for effective computer-based instruction"(p. 395). The construction rather than selection of responses would, in this case, be more appropriate if the desired outcome was the production of acquired skills or knowledge.

Confirmation

Lockee et al. (2004) acknowledged the vagueness of the term confirmation. The differences between feedback, confirmation, and reinforcement are not only philosophical because the implementation of each implies a different relationship between the learner and the instruction. As an example, Lockee et al. noted that Skinner considered confirmation to be positive reinforcement in the operant conditioning process, while others disagreed with this position. The underlining issue is that the terms are sometimes erringly used interchangeably in the literature and this fact might affect research results.

In research they examined, Lockee et al. (2004) found that most studies reported no significant difference. These findings were viewed with caution because the research was labeled as incomplete and lacking information on issues like feedback delay and feedback amount. In addition, some fundamental questions remain unanswered, for example, are signs of progress (correct answers) a significant reinforcer?

Current and Future Research

Research in Educational Technology has shifted from behavioral principles to cognitive and constructivist approaches. This shift has consequently led to a sharp decrease in research on PI. The 1963 PLATO (Programming Logic for Automatic Teaching Operation) project at the University of Illinois is an example of the paradigm shift in research. Originally rooted in the behavioral approach and PI, it has changed to its current more constructivist form (Lockee et al., 2004). Even in its new form, however, PLATO still incorporates behavioral principles, for example, the use of immediate feedback, learner assessment, and performance evaluation.

Although current research is scarce, it does occur. One example of current PI research is Davis et al. (2005). They investigated incremental prompting as a feature of online PI and found that when compared to standard PI and simple prose presentation, using incremental prompting produced significantly better scores on a subsequent applied task that required the learners to write three essays based on the instruction. Davis et al. showed that PI was a viable instructional method reaching higher levels of Bloom's taxonomy, but the results were tempered by the fact that incremental prompting takes a significantly longer time for learners to complete.

In their paper, McDonald et al. (2005) discussed the implications of PI and proposed that certain assumptions led to its decline, and if unaddressed, these assumptions will adversely affect the future of PI:

- Ontological determinism
    - The student's behavior and learning are governed by natural laws
- Materialism
    - Only the overt, measurable behaviors are important
- Social Efficiency
    - The imperative to reduce cost and deliver instructions to wider audiences significantly affected educational practices
- Technological Determinism
    - Technology as the most important force in change significantly affected educational practices.

These assumptions, according to McDonald et al. created an environment where PI materials were rigid and could not be individualized for a particular setting or learner. The prepackaged materials were consequently only useful under specific circumstances, but they were widely distributed and expected to be productive under all conditions. PI thus fell into disfavor because it could not work under all conditions.

If current and future researchers are to learn from PI, they must carefully consider the assumptions to avoid the pitfalls of each. Online learning, which is currently the leading example of instructional technology, assumes both social efficiency and technological determinism (McDonald et al., 2005). It is now frequently designed for the "lowest common denominator" and as cheaply as possible, to the detriment of individualized instruction. To curve this trend, it is important to address these assumptions and adopt creative strategies, for example, using multiple instructional methods within individual instructional units.

McDonald et al. (2005) drew parallels between PI and current forms of instructional technology. It is clear that PI is deeply rooted in the history of instructional and educational technology, and it is reasonable to suggest that the future of both is interconnected. The principles of PI can inform current and future instructional development, but they must be examined and adapted to meet the needs of individual learners. The examination of fundamental assumptions thus becomes paramount. This, however, is not particular to PI; all instructional methods should be carefully examined.

Adaptive Instruction

Park and Lee (2004) defined adaptive instruction as "…educational interventions aimed at effectively accommodating individual differences in students while helping each student develop the knowledge and skills required to learn a task" (p. 651). They presented three ingredients of adaptive instruction: first, the availability of many goals from which to choose; second, the ability to adapt to the learner's initial abilities, then adjust to the particular strengths and weakness of the learner; finally, the ability to help the learner master the instruction then apply that mastery to real-world situations.

Papanikolaou, Grigoriadou, Kornilakis, and Magoulas (2003) in their article on Adaptive Educational Hypermedia systems further clarified the concept of adaptation. They surmised that in order for instruction to be adaptive, the educational environment must make adjustments to accommodate to the learner's need, maintain the appropriate interaction context, and increase the functionality of the educational environment by personalizing the system. They noted two forms of adaptation, adaptivity and adaptability. Adaptivity in their description refers to interactions where the system uses data provided by the learner to modify its controls or functions. Adaptability refers to the ability of the system to support learner initiated and controlled modifications. The basic premise is that adaptation involves changes that are either system determined or learner determined.

Park and Lee (2004) listed three approaches to adaptive instruction:

- Macro-adaptive instructional systems
    - Functions such as instructional goals and delivery system are based on the learner's ability and achievement level
- Aptitude-treatment interactions models
    - Specific instructional strategies are delivered based on specific characteristics of the learner. This model relies on the accurate identification of the most relevant learner traits in selecting the best instructional strategy.
- Micro-adaptive instructional models.
    - Instruction is adapted based on the learner's progress through the instruction. The system continuously monitors the learner's progress and adapts the instruction based on the current performance.

There are advantages and disadvantages to each approach, but for the purposes of the current study, the focus will be on micro-adaptive instructional models. These models are the most relevant because PI fits comfortably within their scope. PI, as previously mentioned, can be developed using branching methods. Branching PI is inherently micro-adaptive in nature because an a priori logic scheme can be used to determine the behavior of the adaptation, that is, the branch. The constant monitoring of the learner's performance, which is a trademark of PI, easily satisfies the requirement that micro-adaptive instruction diagnoses the learner's current performance and adapts the instruction based on that diagnosis.

Micro-Adaptive Instructional Models

Micro-adaptive instructional systems rely on the ongoing process of learner diagnosis to determine an optimal instructional path (Park & Lee, 2004). Although the fundamental issue of continuous diagnosis is prevalent in historical attempts at this model, several schools of thought have emerged. Park and Lee (2004) discussed several views noting subtle differences between the implementations of micro-adaptive models. For example, perspectives that adapt the sequence of instructions can be compared with those that adapt the volume of instructional content delivered. The implementations have similarities, but they are conceptually different and have important implications.

Several micro-adaptive models have been developed and researched over years. Park and Lee (2004) discussed several models, but the Trajectory and the Bayesian models are of particular importance. Specific features from these models can be adapted or discarded, while others can be modified to produce a tailor-made foundation for the current study.

The Trajectory model as described by Park and Lee (2004) uses numerous learner traits to predict the next step in instructional sequences. In theory, the Trajectory model uses the learner traits (group variables) to determine an optimal instructional approach, although individual learner attributes are included during diagnosis and prescription of instructional steps. This model is not natively compatible with PI. Firstly, the Trajectory model accommodates group properties and individual internal states descriptions to predict at minimum the sequence of instructions. Secondly, the large numbers of learners necessary to generate an effective predictive database is also a limiting factor in the implementation of this model. Finally, this model uses only a few variables because accounting for a large number of variables is developmentally unrealistic. Those three examples taken in isolation make the case that this model is irrelevant to PI, but it was based on this model that Ross and Rakow developed an adaptive system that can be modified for use with PI. The Ross and Rakow model that is cited in Park and Lee (2004, p. 664) uses the Trajectory model to determine the best strategy for selecting the appropriate number of examples in a given instructional sequence.

The core tenets of the Ross and Rakow model can be used within the PI context because it is possible to add examples to an instructional sequence without altering the overall direction and linearity of the content. The current study adapts the underlying Ross and Rakow principles and delivers examples and non-examples based on an evaluation of the learner’s current response. The overall sequence of instruction is not altered, except that examples and non-examples are inserted into the instructional sequence as necessary.

The Bayesian Model presented in Park and Lee (2004) uses a two-step approach for adapting instruction. First, a pre-instructional measure is delivered and its results are used to determine the treatment that the learner will experience. Second, the instruction is adjusted based on the continuous monitoring of the learner's performance. Bayes theorem of conditional probability is then used to predict mastery level based on the pre-instructional measure and the on-task performance.

Like the Trajectory model, the practical implementation of the Bayesian model requires a large sample before predictions are reliable. The sample size requirement is a serious limiting factor, and in addition, generating a reliable distribution based on prior learners' scores and historical data is not a trivial task. The complexities of Bayes Theorem need not be applied if the simpler algorithms can be used to accomplish the same task.

The key concept from the Bayesian Model that is usable within the scope of the current study is the use of a pre-instructional measure as a means of determining a starting position for instruction. The only caveat to delivering a pre-instructional measure is that the purpose of the measure must be made clear in advance. Papanikolaou et al. (2003) justified the use of a self-report pre-instructional measure as the basis for their INSPIRE system. In this system, learners reported their learning style preference and the subsequent instruction were based upon that report. An interesting observation is that at any point during the treatment, learners could change their learning style preference. While it is puzzling that the system was designed to allow learners to change their learner style preference in the middle of instruction, it can be argued that the inclusion of this function indicates that the developers were confident in the validity and dependability of self-reported data, and also that they were comfortable with premising their design on a purely theoretical knowledge base – learning styles.

Self-reported data can be very useful but it should not be a fundamental and critical part of the instruction. The argument can be made that most learners are not aware of how they learn best so they might not be in the best position to comment on their own learning style. Gilbert and Han (1999) developed an adaptive system, ARTHUR, that was sensitive to learning styles without using self-reports. In their system, the learners were exposed to instruction presented in Audio, Video and Text formats. Learners' performance on each item from each format would form the basis for the branching of future learners, thus a robust classification matching learner to learning style would evolve as more and more cases populate the system. Clearly the need for a large user population is a disadvantage of the system, but the accommodation of individual learning styles without the use of self-reports is feature worthy of comment.

The Trajectory and Bayesian Models provide key features that can be applied to the current study. The idea of a pre-instructional measure to determine starting position and constant monitoring and adaptation to refine instruction are complementary concepts that can coexist in any well-designed system. This idea is supported by Stoyanov and Kirchner (2004) in their statement:

> An adaptive e-learning environment should propose opportunities for pre-assessment of different personal constructs, for example, through tests or check-lists for knowledge, and questionnaires for learning styles and learning locus of control. Based on a personal profile, the user receives suggestions of what and how to study. This pre-specified adaptation could coexist with dynamically adapting instruction while tracking learner behavior. Most current theory and

> practice in e-learning suggest either pre-assessment or monitoring adaptation, but rarely a combination. (p. 50)

The current research greatly benefits from the clarity that a foundation in PI affords. As discussed earlier, the programming of instruction is paramount in PI. Given this fact, it is clear that the ability to alter the sequence of instruction (frames) would not be of any value because shaping cannot occur in the presence of (possibly) disjoint frames. The sequencing of frames, which is the essence of programming, would be broken if the system were to either support the learner's ability to alter the instruction sequence – adaptability, or determine the instruction sequence via an algorithm – adaptivity. In this case, the best solution, conforming to the requirements of PI, would a system that adapts the volume of instruction that it presents to the learner. Within the PI construct, this can be implemented in terms of how many examples and non-examples the learner experiences, and the conditions under which the adaptivity occurs.

Ray (1995b) described a system called MediaMatrix that facilitated adaptive instruction. In his system, tutorials can be built such that learners are exposed to alternate forms of the current instruction based on their current performance. This system was built on the behavioral principles of shaping and successive approximation, thus the branching to alternate option would not violate the behavioral principles of a systematic linear progression to the terminal behavior. It is interesting to note that the system does not itself shape performance, but instead, it provides the environment through which a programmer could produce instruction that shapes appropriate behavior.

In addition to dynamically branching to alternate instruction, Ray (1995a) also described the use of pop quizzes as a means of branching control. In this case, the learner is presented with a pop quiz selected randomly from any previously completed topic. An incorrect response branches the learner to a lower level programming density, assuming that the learner has not mastered the concept and needs additional help. The lower level programming density refers to instruction containing more prompts and extra stimuli for the benefit of the learner. If the learner's performance does not improve as the levels decrease, thus increasing the programming density, the learner is presented with alternate forms of the instruction. The current study did not implement the pop quiz mechanism described in Ray (1995a), but it implemented branching to alternate forms of instruction, where the presentation of additional examples or non-examples can be described as an alternate form of instruction.

Research on adaptive instruction remains unclear. Park and Lee (2004) reviewed studies where adaptive instruction was demonstrated to be effective, and also where no significant difference existed. This fact is highlighted by Song and Keller (2001) where they investigated motivationally adaptive instruction. They found both significant and non-significance among variables, for example, motivationally adaptive instruction, motivationally non-adaptive, and minimal non-adaptive instruction. In certain circumstances, adaptive instruction was superior but in others, no significant difference was observed. The study concluded that adaptive instruction is feasible and can be effective, but they conceded that additional research would be needed to verify their inferences as to why the differences, or lack thereof, occurred.

The inconclusiveness of the research does not undermine the importance of adaptive instruction. On the contrary, adaptive instruction is almost a requirement in any good instruction system. Park and Lee (2004) suggested that "A central and persisting issue in educational technology is the provision of instructional environments and conditions that can comply with individually different educational goals and learning abilities" (p. 1) . The literature emphasizes that although adaptive instruction is viewed as important, the tools and methods required to successfully implement effective systems are still in their infancy (Park & Lee, 2004).

Intelligent Tutoring Systems

Intelligent Tutoring Systems (ITS) must "…(a) accurately diagnose students' knowledge structures, skills and/or styles using principles, rather than preprogrammed responses, to decide what to do next; and then (b) adapt instruction accordingly" (Shute & Psotka, 1996, p. 576). It is designed to mechanically foster a dialogue resembling the interaction between a teacher and a learner (Park & Lee, 2004), and this interaction is mediated by complex algorithms and artificial intelligence (AI) methods for describing the teaching and learning process.

ITS are designed to expertly adjust to the learner's performance thus maximizing the instructional value of the content. The emphasis on expert behavior is highlighted by Shute and Psotka (1996) when they suggested that the system must behave intelligently, not actually be intelligent as in human intelligence. This inherent complexity adds to the flexibility of an ITS. Pre-determined branching rules that are normally a part of adaptive systems, are replaced by more dynamic methods that seek to abstractly represent knowledge.

ITS are conceptually very powerful and promising, but their complexity is a deterrent to practical application. AI methods of representing knowledge are not only difficult to invent, but "…using fancy programming techniques may be like using a shotgun to kill a fly. If a drill-and-practice environment is all that is required to attain a particular instructional goal, then that's what should be used" (Shute & Psotka, 1996, p. 571). This review does not consider ITS at any depth because given the nature of the current study, complex algorithms and AI methods might be considered shotguns.

Simulations

What is a Simulation?

There is considerable debate over the definition of simulations (Alessi & Trollip, 2001). Current perspectives view simulations as complex visual environments involving many user accessible variables. Gredler (2004) characterized simulations as open-ended, evolving, and having many interacting variables. She defined a simulation as "an evolving case study of a particular social or physical reality in which the participants take on bona fide roles with well-defined responsibilities and constraints" (p. 571). She subsequently presented several important attributes that delineate a simulation from other forms of instruction:

1. A complex real-world situation modeled to an adequately degree of fidelity.
2. A defined role for participants, with constraints.
3. A data-rich environment.
4. Feedback in the form of changes to the modeled situation.

In her estimation, solving well-defined problems does not constitute a simulation; instead a simulation is an ill-defined problem with many variables, and many possible solution paths. She used the term "Deep Structure" (Gredler, 2004, p. 573) to describe a central feature of simulations. Deep structure does not only suggest multiple solution paths, but includes the fact that a participant must be a part of the experience such that each action has valid consequences associated with that action.

Gredler (2004) viewed complexity as a necessary trait of simulations, and while this is true in some cases, simulations purposed at learning need not contain many variables and multiple solution paths. Alessi and Trollip (2001) supported the broader characterization of simulations when they not only described various different types of simulations but defined educational simulations as “a *model* of some phenomenon or activity that users learn about through interaction with the simulation” (p. 213). They also noted that “A simulation doesn’t just replicate a phenomenon; it also simplifies it by omitting, changing, or adding details or features” (p. 214). Alessi and Trollip’s definition is significant in that it does not comment on the complexity or the degree of fidelity required before an application can be called a simulation. This difference is not trivial because the current study assumes that the prevailing perception of simulations as complex environment akin to flight simulators is too narrow and thus lacks the ability to support applications geared for educational purposes.

The idea of complexity also affects the fidelity of simulations. High degrees of fidelity are not necessarily a prerequisite for simulations. The relationship between fidelity and performance is not linear (Alessi & Trollip, 2001), thus more visually realistic simulations are not necessarily better for learners. Couture (2004) found that some high fidelity characteristics of the simulated environment resulted in higher credibility but other high fidelity characteristics had the opposite effect. Couture (2004) attributed the results to particular learner characteristics.

Games, Virtual Reality and Microworlds

One reason the definition and character of simulations remain unclear is the fact that similar but different technologies are often defined as and used synonymously with simulations. Games, Virtual Reality and Microworlds can all have simulation components, but they differ on at least an abstract level from a simulation.

A game can contain simulation elements and likewise a simulation can contain gaming components. Educational games are defined as competitive exercises where the goal is to win (Gredler, 2004). The key features include competition with other players or the environment, reinforcement for correct actions in the form of advancement, and actions that are governed by rules that may be imaginative.

Gredler (2004) listed four reasons for using educational games: practice, identification of weaknesses, revision, and development of new knowledge and/or skills. The primary difference between an education game and a simulation is competition. Games use competition primarily as a motivator, but this is absent from simulations. There are however educational simulation games, that combine both concepts in an effort to facilitate learning. While competition is the dominant aspect of simulation games, for example SimCity, the simulation also provides an opportunity to learn the underlying model.

McLellan (2004) defined Virtual Realities as “as a class of computer-controlled multi-sensory communication technologies that allow more intuitive interactions with data and involve human senses in new ways” (p. 461). Virtual reality applications allow very high degrees of interaction, allowing the user to have experiences from many viewpoints using multiple senses. The enhanced interaction is usually experienced using specialized equipment such as head mounted monitors and gloves that provide tactile feedback.

Although virtual reality has different classifications depending on the author involved, a few common threads exist amongst the competing views. For example, virtual realties model environments that may be physical or abstract, and the user is able to access and manipulate aspects of the model and receive action-dependent feedback (Reiber, 2004).

Microworlds are environments where learners explore and build knowledge. Reiber (2004) explained that the learner must “get it” (p. 587) before an environment is considered a microworld, thus the definition of microworlds is tied to their function. According to Reiber, microworlds are: domain specific, facilitate the easy acquisition of the domain-required skills, are derived from a constructivist perspective, and afford immersive activities that are intrinsically motivating.

Reiber (2004) distinguished microworlds from simulations and other tools by discriminating between using models and building models. Simulations allow learners to work with a pre-existing model, thus manipulating ready-made tools presented on the interface. Contrary to simulations, microworlds allow learners to build the tools they will use. In this case, the learner is not limited to the variables and parameters described by the interface.

Games, virtual reality and microworlds each exist along a continuum but it is not immediately clear where one begins and the others end. Although they can be defined and described separately, each can at some level subsume or be subsumed by simulations. This reality does not undermine the credibility or usefulness of simulations, instead it highlights the fact that good instruction can and sometimes should contain various methodologies.

Types of Simulations

The cited works, Gredler (2004) and Alessi and Trollip (2001), have slightly differing definitions of simulations. Although there is considerable overlap, the differences in the conceptualization of what exactly is a simulation also leads to similar but different classifications of simulations. Gredler (2004) defined two broad types of simulations. First, experiential simulations provide a means for learners to interact with systems that may be too costly or dangerous for real-world experimentation. It is described as a "social microcosm" (p. 573) where individuals have different responsibilities in the complex evolving scenario. This type is further divided into:

- Social process
    - Consequences are embedded in the scenario.
- Diagnostic
    - Consequences are based on optimal responses.
- Data management
    - Consequences are embedded in the relationship between associated variables in the system

Second, symbolic simulations are models of specific functions or phenomena of another system. In this case, the learners play the role of researchers investigating an external system. The type is further subdivided into:

- Laboratory-research simulations
    - Individuals interact with a complex system to solve a problem.

- System simulations
  - Individuals interact with a complex system to diagnose a problem with that system.

Alessi and Trollip (2001) categorized simulations in a significantly simpler manner. They used two categories, “About something” simulations and “How to do something” (p. 214) simulations. These were further subdivided as follows:

“About something” simulations

- Physical
  - Learners interact with the representation of a physical object or phenomenon.
- Iterative
  - Learners interact with the representation of a physical object or phenomenon, but do so at discrete points where they vary inputs and observe the new simulated result.

“How to do something” simulations

- Procedural
  - Learners interact with a system designed to teach a sequence of steps to accomplish a task.
- Situational
  - Attitudes of people or organizations under differing circumstances are investigated.

The current study is best described in terms of the simple and clear definition proposed by Alessi and Trollip (2001). The instructional application developed in the current study can be described as a procedural simulation, but both Alessi and Trollip (2001) and Gredler (2004) admitted that the categories are not mutually exclusive, hence considerable overlap is possible. The current study, although procedural, contains components that are physical simulations and iterative simulations. This is understandable when the nature of the study is considered; learning *how* to graph is only successful if the learner also knows *about* graphs.

Advantages and Disadvantages of Simulations

Simulation as a method of instruction and learning is predicated on the assumption that it is inherently interesting and motivating. While disadvantages are freely acknowledged, the purported advantages are the driving force behind the continued interest in simulations. While simulations have obvious advantages, such as cost and safety, over real world learning environments, they are described as having particular benefits over other instructional methodologies. Alessi and Trollip (2001) listed four main advantages of simulations have: motivation, transfer of learning, efficiency, and flexibility:

- Motivation
    - Most simulations contain graphical elements that are at minimum reinforcing for learners. Coupled with active participation and feedback, a graphically rich simulation presents an environment where learners are more likely to enjoy the learning process.
- Transfer of Learning
    - Simulations tend to facilitate transfer because of their ability to engage various solution paths for a problem.

- Efficiency
    - The initial learning efficiency and effectiveness can be reduced using well designed simulations.
- Flexibility
    - Simulations can be designed to present various components of instruction. They may present information, guide learning, facilitate practice, or assess knowledge. In addition, simulations are unique in that they are applicable across learning philosophies.

Gredler (2004) listed three other key advantages to simulations and asserted that

1. They can reveal learner misconceptions about the content.
2. Simulations present complex environments that bridge the gap between classroom and the real world
3. They can provide insight into the problem-solving strategies that learners use.

Despite their appeal and significant advantages, simulations are by no means the perfect methodology (Alessi & Trollip, 2001; Gredler, 2004). Several disadvantages were identified

- Simulations can be prohibitively costly to produce. The financial and temporal costs involved in developing a simulation might not be worth the expected educational benefits.
- The benefits of a simulation might be overly dependent on prerequisite knowledge, thus learners might need prior instructions before they can engage in open-ended learning.
- Learners may experience cognitive overload in complex environments.

Other general disadvantages include the inability of simulations to accommodate sub-goals, empathy, or differing learner levels. These criticisms, however, can be leveled against any instructional software that is not specifically designed to deal with those specific issues.

Are Simulations Effective?

The effectiveness of simulations remains unclear. Some studies demonstrate significant differences in favor of simulation while others report no significant differences. Most studies, however, acknowledge the potential benefits of simulations and discuss specific ways to help maximize the learning experience. The thematically similar studies chosen for this section not only reflect the uncertainty of the effectiveness of simulations in general, but the controversy that exists within the selected content areas.

Simulations have been studied using many different content areas, but science seems to be most suitable for simulations (Lee, 1999). Although science is particularly suited for simulations, the results are not always consistent.

Steinberg (2000) investigated the differences between a computer simulated interactive learning environment and a traditional pen-and-pencil interactive environment. He used calculus-based physics as the primary content area, focusing specifically on air resistance. Of the three classes involved in this study, two classes (n=79 and n= 67) were administered the simulation based content and the other class (n=83) experienced a pen-and-pencil version of the same content. Steinberg found that although there were differences in the learners' approach to learning, there was no significant difference in posttest scores. Although anecdotal observations regarding classroom events and study artifacts were mentioned, no rationale as to why the study was conducted after seven weeks, or the way the groups were divided, was provided.

Campbell, Bourne, Mosterman, and Brodersen (2002) investigated the effectiveness of an electronic laboratory simulator versus physical laboratory exercises. The content was primarily focused on building electronic circuits. The study had two groups, the combined laboratory group (n=22) used electronic simulated labs then two subsequent physical labs for practice. The physical laboratory group (n=18) did not use any simulations. Campbell et al. found no significant differences between the groups on a pretest measure, but found significant differences on the written posttest. There was no significant difference in the time it took to complete the final physical laboratory that was a group task, thus the combined group did at least as good as the physical group.

Lane and Tang (2000) investigated the use of simulation within a statistics context. Their study involved a group that received simulation-based instructions, a group that received textbook-based instruction, and a no-treatment group. They found that the simulation group performed significantly better on the posttest and were able to better apply statistical concepts to real world situations. Lane and Tang admitted that the implemented simulation might not be considered a simulation by some, for example (Gredler, 2004), because it was merely a video presentation simulating the concepts. According to the authors, “it is likely that the advantage of simulation in the present experiment would have been greater if learners had had the opportunity to interact with the simulation” (Lane & Tang, 2000, p. 351).

The use of simulations in statistics was also examined by Mills (2004). This study considered whether or not computer simulated methods enhanced learners’ comprehension of abstract statistics concepts. The two instructional units used for the study can be described as interactive versus non-interactive, thus it is unclear if factors such as motivation can be considered confounding. Although the sample size was small, Mills concluded that learners can learn abstract statistical concepts using simulations.

Spelman (2002) examined simulations under unique conditions. Unlike the previously discussed studies, Spelman examined the GLOBECORP business simulation and its effect on writing proficiency over the course of a semester. Based on the data gathered, Spelman found that the learners who used the simulation did at least as well as the learners who used the traditional format based on posttest results and a writing proficiency test. The learners who experienced the simulation reported significantly less anxiety and significantly greater satisfaction with the instruction. These results were used to assert that "instructors who wish to enliven their classrooms by changing to approaches that include simulation should do so with confidence" (Spelman, 2002, p. 393)

de Jong and van Joolingen (1998) extensively reviewed the literature on discovery learning with simulations, focusing on learner problems. Based on a selected subset of their reviewed literature, they concluded that "The general conclusion that emerges from these studies is that there is no clear and univocal outcome in favor of simulations" (p. 182). The review did not entirely focus on whether or not simulations were effective, but they found significant findings both in favor of and against simulations. According to the authors, if a theory of discovery learning using simulations is to be developed, more research should be done in identifying the problems learners have in this environment, and in developing tools to address those problems.

A meta-analysis conducted by Lee (1999) reviewed literature based on two types of simulations: pure and hybrid, and two instructional modes: presentation and practice. Lee concluded that

1. Within the presentation mode, the hybrid simulation is much more effective than the pure simulation.
2. Simulations are almost equally effective for both the presentation and the practice mode if hybrid simulation is used.
3. Specific guidance in simulations seems to help learners to perform better.
4. When learners learn in the presentation mode with the pure simulation, they showed a negative attitude toward simulation.
5. Even in the practice mode, learners showed very little preference to simulations.
6. Science seems to be a subject fit for simulation type of learning.

Lee (1999) cautioned that no definite conclusions should be attempted due to the small number of studies in the meta-analysis and the possible existence of other well-designed studies.

Although the effects of simulation are inconsistent, the reviewed studies reveal common threads. Simulations were reportedly more motivating and at least as good as traditional instruction, thus its inclusion in instruction would not be harmful. Although the need for more research was emphasized, simulations were thought to be excellent if they were a compliment to other forms of instruction. This is most clearly demonstrated by Wiesner and Lan (2004) when they found that of 12 oral presentation teams, 9 recommended a combination of simulation and physical labs, 3 recommended only physical labs, and none recommended only simulations.

Why use a simulation

"Simulation is one of the few methodologies embraced equally by behavioral versus cognitive psychologists, and by instructivist versus constructivist educators. Simulations can be designed in accordance with any of these approaches" (Alessi & Trollip, 2001, p. 231).

The instructional application developed in the current study focused on graphing instruction, and as such, the nature of the task makes it ideally suited for a simulation. Although the effects of simulations on learning are inconsistent, the research has shown that under the right conditions, simulations are a viable option and can return significant learning gains. Combining direct instruction, discrete content instruction, and simulation exercises might be the conditions for significant learner gains.

The Design of Instruction

The design of instruction has been described as an art, a science, a technology, and a craft (Wilson, 2005). Although there are many ways to produce instruction, there are generally accepted methods to do it effectively. This section will focus on the systematic design of instruction, thus concentrating on the viewpoint: the design of instruction as a science.

What are Instruction and Design?

Instruction is generally defined within the scope of learning. Alessi and Trollip (2001) defined instruction as "the *creation and use of environments in which learning is facilitated*" (p. 7). Gagne et al. (2005) proposed that instruction is a "set of events embedded in purposeful activities that facilitate learning" (p. 1). Behaviorally, the terms *environments* and *events*, which were used in the previous definitions, can be viewed as arrangements of contingencies. This perspective is important because it not only accounts for the environment and events themselves, but also the interaction between different properties, for example, sequences or lengths.

P. L. Smith and Tillman (2005) supported the behavioral perspective and proposed that instruction is "the intentional arrangement of experiences, leading to learners acquiring particular capabilities" (p. 5). They further clarified their definition by discriminating between instruction, education, training, and teaching. Education, in their context, broadly describes all experiences where people learn, including experiences that are non-intentional, unplanned, or informal. Training refers to instructional experiences where the primary focus is the acquisition of a specific skill that will be immediately useful. The immediacy of the skill's application is the defining feature of training. Teaching refers to learning experiences that are mediated by humans.

Although instruction, education, training, and teaching are often used interchangeably, they have subtle differences. These distinctions provide focus and allow the current study to be validly grounded in the framework of *instruction*.

Smith and Tillman (2005) defined design as "an activity or process that people engage in that improves the quality of their subsequent creations" (p. 6). Although design and planning are sometimes used synonymously, the design process involves a higher degree of preparation, precision, and expertise. The implication is that the design process should move the entire production process closer to the realm of scientific. Gagne et al. (2005) adopted six assumptions regarding the process of design:

1. Instructional design is focused on the process of learning rather than the process of teaching, and the aim is intentional learning as opposed to accidental learning.

2. The learning process is complex and involves many variables that may be related.
3. Instructional design models can be applied at different levels, for example, at the lesson module level or at the course level.
4. The design process is iterative and successive refinements are required to produce effective instruction.
5. The design process is itself comprised of sub-processes.
6. Different outcomes call for different designs.

Instructional Systems Design

Instructional Systems Design (ISD) and its relation to Instructional Design (ID) remain unclear in the literature. The lack of clarity begins with the acronym ISD. The term Instructional Systems *Development* (Dick et al., 2005; Wilson, 2005) and Instructional Systems *Design* (Gagne et al., 2005; P. L. Smith & Tillman, 2005) are both used for ISD. The current study will use Instructional Systems *Design*, supporting the definition “the process for creating instructional systems” (Gagne et al., 2005, p. 18) where instructional systems include resources and procedures used to facilitate learning.

The relationship between ISD and ID differs depending on the author. Morrison et al. (2004) defined ID as “The systematic method of implementing the instructional design process” (p. 5). In that definition, ID is systematic thus negating the need for a separate ISD. Other authors distinguish between the two, viewing ID and an overarching term and ISD as the systematic and scientific way to do ID (Dick et al., 2005; Gagne et al., 2005; Wilson, 2005). The current study will use ID as an umbrella term, and ISD as a systematic implementation of ID.

The final relevant issue regarding the clarity of ISD is its description. Some authors describe ISD as a model (Wilson, 2005), others describe it as a process (Gagne et al., 2005), and yet others describe it as both model and process (Dick et al., 2005). The issue is not trivial because the current study has different components that must be clearly defined before they can be validly used. The current study will use ISD as a term to describe a general process. The transition will then be made to a specific exemplar model describing that process.

A Systematic Approach

For more than 40 years, ISD has been taught as the primary framework for instructional design. Although it has been in use for many years, an empirically based body of research supporting ISD does not exist. Wilson (2005) suggested that along with the fact that it is difficult to scientifically test comprehensive processes, the axiomatic status of ISD within the field has led to the current scenario where ISD is accepted without question. Dick et al. (2005) suggested that ISD is valid because the component parts of ISD are based on validly researched learning principles. In essence if the parts are valid, the whole must also be valid.

There are many logical and conspicuous reasons why a systematic approach to the design of instruction is beneficial. Smith and Tillman (2005) provided seven advantages of a systematic approach:

1. Encourages advocacy of the learner by making the learner focus of the instruction.
2. Supports effective, efficient, and appealing instruction.
3. Supports coordination among designer, developers, and those who will implement

the instruction.

4. Facilitates diffusion, dissemination, and adoption because the products are more likely to be practical and easily duplicated.
5. Supports development for alternate forms or delivery systems. This is particularly useful because modularity facilitates future development.
6. Facilitates congruence and consistency among objectives, activities, and assessment.
7. Provides a systematic framework for dealing with issue, for example, learning problems.

The systematic design of instruction, though very useful, has several disadvantages. Smith and Tillman (2005) listed some of the limitation of ISD, placing particular emphasis on the inability of ISD to facilitate goal-free learning, or situations where learning goal cannot be determined in advance.

Gordon and Zemke (2000) not only considered the limitations of ISD, but they questioned the usefulness and relevance of ISD. They spoke to six experts and compiled the criticisms into four major categories:

- They argued that ISD is too slow and clumsy, citing the fact that real business cannot afford the time investment needed to complete each component of ISD.
- There’s no “there” there. This is a criticism of the rigidity of ISD and the pursuit to make it scientific.
- They positioned that if ISD is used as directed, it produces bad solutions. They supported this criticism by citing designers’ preoccupation with learning styles and best practices, neglecting the purpose of the training.

- According to the authors, ISD clings to the world view of "weak learners and all-knowing experts". This wrong view option results in the production of training that caters to the least common denominator.

These criticisms are based on ISD as a framework for training. This review has previously discriminated between training and instruction, and although they are not necessarily the same, the criticism appears to be valid in both areas.

Critics of ISD probe the failures of ISD trying to determine if the process, the practice, or the definition is flawed (Zemke & Rossett, 2002). Proponents of ISD continue to implement new techniques within the ISD framework and for them, ISD remains significant and relevant. "Suggesting that the ISD process is no longer relevant to 21st-century training is the equivalent to suggesting that engineers forget about data flow and process diagrams" (R. C. Clark, 2002, p. 9). The reviewed literature, however segmented on many issues, concluded that ISD is not perfect, but it can be very useful if properly applied.

ADDIE

The Analysis, Design, Development, Implementation, and Evaluation (ADDIE) model describes the ISD process and forms a framework for effective instructional design. While the exact origin is unknown, the model has served as both a stand-alone framework, and as a foundation for other more specialized models (Molenda, 2003).

Gagne et al. (2005, p. 21) provided a pictorial view (see Figure 3) outlining the interactions among the various phases:

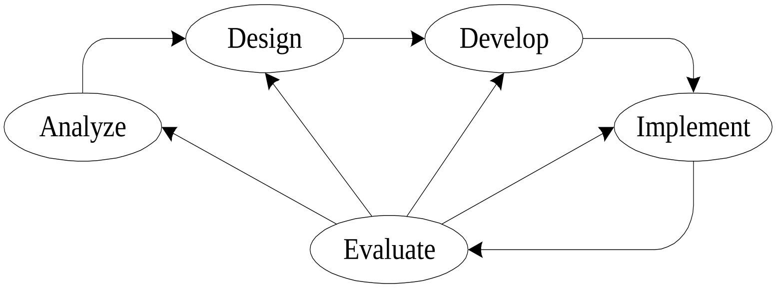


Figure 3. Gagne's ADDIE model.

The layout of the model implies an output/input relationship where the output of one phase becomes the input for the next phase. This linear progression leads to organized development and provides the opportunity for accountability at each phase. Accountability is possible because each phase can be evaluated and repeated if necessary.

The ADDIE model can be viewed as a generic prototypical representation of the ISD process. It does not facilitate the development of all possible types of instruction, but all other ISD models can be reduced to at least a subset of the ADDIE phases (Gagne et al., 2005).

The ADDIE Phases

Each phase of the ADDIE model is purposed at accomplishing defined goals. There is no generally accepted list of sub-components for each phase, thus the specific sub-components of each phase can be viewed as a set of best practices proposed by individual authors. Gagne et al. (2005) provided a summary that formed the foundation for the current study. This summary was selected because it is both comprehensive and consistent with other summaries reviewed.

Gagne et al. (2005, p. 22) summarized the five ADDIE phases and sub-components as follows:

1. Analysis
    a. First determine the needs for which instruction is the solution.
    b. Conduct an instructional analysis to determine the target cognitive, affective, and motor skill goals for the course.
    c. Determine what skills the entering learners are expected to have, and which will impact learning in the course.
    d. Analyze the time available and how much might be accomplished in that period of time. Some authors also recommend an analysis of the context and the resources available.
2. Design
    a. Translate course goals into overall performance outcomes, and major objectives for each unit of the course.
    b. Determine the instructional topics or units to be covered, and how much time will be spent on each.
    c. Sequence the units with regard to the course objectives.
    d. Flesh out the units of instruction, identifying the major objectives to be achieved during each unit.
    e. Define lessons and learning activities for each unit.
    f. Develop specifications for assessment of what students have learned.
3. Development
    a. Make decisions regarding the types of learning activities and materials.
    b. Prepare draft materials and/or activities.
    c. Try out materials and activities with target audience members.

    d. Revise, refine, and produce materials and activities.
    e. Produce instructor training or adjunct materials.

4. Implement
    a. Market materials for adoption by teachers or students.
    b. Provide help or support as needed.
5. Evaluate
    a. Implement plans for student evaluation.
    b. Implement plans for program evaluation.
    c. Implement plans for course maintenance and revision.

The current study could not implement each sub-component as they are, instead each phase and sub-component was assessed and modified for the current task.

Why use ADDIE?

The unique feature of ADDIE, that it is both specific and general, is the source of both its greatest praise and criticism. The adaptability of the model makes it perfect for novice designers and environments where goals and objectives are not initially fully defined, as is the case of the current study. While other ADDIE-based models, for example Dick et al. (2005), appear complete, the generic ADDIE model provides sufficient leeway affording a basic development structure and simultaneously facilitating experimentation and encouraging adaptation.

Graphing

Teaching learners to represent experimental data in graphical form and interpreting already existing data has traditionally been a difficult task. Techniques such as task modeling with instructor guidance have generally been used to get learners to apply rules and generate propositions that eventually lead to the completion of tasks.

Graphs and the process of graphing are important not only in scientific dialogue where data relationships are pursued, but also within the lay context where consumers of data must discriminate between available choices. Monteiro and Ainley (2002) posited that "As a data handling activity, graphing might be conceptualized as a process by which people can establish relationships between data, and infer information through the construction and interpretation of graphs" (p. 61). Making valued judgments in the presence of erroneous or purposefully skewed data is a requirement of the current consumer of data (Gal, 2002).

In summarizing Bruno Latour's Essay "Drawing Things Together", L. D. Smith, Best, Stubbs, Johnston, and Archibald (2000) presented the five features of graphs that make them powerful and persuasive:

> First, they are able to transcend scales of time and place, rendering invisible phenomena (such as quarks, ion pumps, gross national products) into easily apprehended icons. Second, they are manipulable, and can be superimposed and compared in ways that lead to seeing new connections between seemingly unrelated phenomena, discerning similarities between events vastly separated in time and place, and assessing the congruence of empirical and theoretical curves. As such, they encourage the sort of abstraction from detail to generalities that is characteristic of theoretical science. Third, graphs are 'mobile' or transportable: they can be carried from laboratory to laboratory, or from laboratories to scientific conferences, or from research sites to sites of application. Fourth, they are

> 'immutable', both in the sense of fixing the flux of phenomena – and thus stabilizing what may be only ephemeral in nature or the laboratory – and in the sense of remaining stable as they are transported across contexts. Fifth, as 'immutable mobiles', graphs can be enlisted in the task of persuading scientists in competing camps of the validity of one's evidence. As Latour puts it, a well-constructed graph raises the cost of dissenting from one's own favoured viewpoint, forcing scientific adversaries to muster their own evidence in the form of even better graphs. To the extent that scientists are able to mobilize consensus on data and evidence, it is through competition and negotiation over graphical representations (hence Latour's motto that 'inscriptions allow *conscription*'). The centrality and pervasiveness of graphs in science led Latour to conclude that scientists exhibit a 'graphical obsession', and to suggest that, in fact, *the use of graphs is what distinguishes science from nonscience*.[15] Others who analyze the representational practices of scientists share Latour's conviction that graphical displays of data play a central rather than peripheral rôle in the process of constructing and communicating scientific knowledge.[16] (p. 75)

Although emphasis is placed on the interpretation of graphs, the experiential factors involved in graphing requires that graphing be viewed as a collection of complex tasks that cannot exist in a vacuum (Brasseur, 1999). It is important to understand that interpreting graphs requires a certain skill set that is most easily obtained via the process of collecting, parsing, and presenting the data. The process of creating graphs thus affects the ability to interpret graphs.

The literature on graphing confirms that the task of graphing is complex and consequently more difficult than traditionally believed (Brasseur, 2001). Although Roth and McGinn (1997) suggested that learner performance may differ by age or other characteristics, in general graphing is understandably paired with general mathematics and science, thus assuming a supposed inherently high level of difficulty. This phenomenon has been confirmed by the researcher during preliminary course improvement research where it was observed that learners performed poorly on graphing tasks even after they had successfully completed graphing instruction. Although the learners could discriminate between good and bad graphing, and to some degree interpret sample graphs, learners failed to independently demonstrate mastery of basic graph production skills. Learners also reacted negatively to the task, questioning the value of actually creating graphs. This result supports Barton (1997) when he suggested that "At this level the difficulties of producing manual plots is not as significant" (p. 70). On the surface it would appear that analyzing a graph is only loosely connected to the ability to create a graph. The preliminary course improvement research concluded that interactive instruction of any sort should contain required graphing responses and provide immediate correction and confirmation, thus aiding learners in developing their graphing skills. Although the preliminary data are mostly anecdotal, it is consistent with the overall picture that the literature paints: graphing is a complex task whose worth and value erringly differs depending on the viewer's lens.

The issue of graphing has both practical and philosophical roots. Practically, the issue is whether or not graphing and similar tasks can be taught without requiring learners to actually produce good, accurate graphs. Younger learners are often given sheets of graph paper that they then use to produce appropriate figures, all under the watchful eye of the teacher. Thus, they have a "sample work space" at their desks where they can actively practice the graphing behavior that is required. In more advanced classes, older learners rarely practice the task of graphing, but are instead expected to know how to graph. In most cases, the teacher or textbook presents a graph and learners can optionally practice on their own if they wish to reproduce the graph. This method is very troublesome in that it makes several assumptions. It assumes that learners have already mastered the “art” of graphing, unconcerned with the notion that not all graphing is the same. It also minimizes the relevance of producing graphs by placing emphasis on the evaluation and interpretation of graphs. While it is important to interpret graphs correctly, the production of accurate graphs cannot be overlooked. More importantly, the production of graphs and the interpretation of graphs are not necessarily mutually exclusive events (Brasseur, 2001). This is especially true in the education environment where teachers can and should analyze the learner and the class such that behavior patterns can be detected and addressed.

Philosophically, the task of graphing is very interesting because of what it represents. L. D. Smith et al. (2000) cited Bruno Latour’s essay “Drawing Things Together” and emphasized the “centrality and pervasiveness of graphs” when they concluded that “graphs constitute the lifeblood of science” (p. 75). This scientific perspective on the importance of graphing is further complicated when the social aspects of graphing are considered. Roth and McGinn (1997) investigated graphs as inscription and conscription devices, placing graphs as mediators of messages and meaning. They presented different lenses though which graphing can be perceived and it becomes clear that the creation and manipulation of graphs are tied to personal prejudices. As noted in Brasseur (1999), “Thus, context and experience with a culture is a key factor for graph designers and interpreters and is just as important as their perceptual skills and their understanding of the technology” (p. 11).

Given the complexity of graphing, the question becomes whether or not graphing is a mental or a task rooted in practice (Roth & McGinn, 1997). The implications are serious, ranging from expectations to assessment. An obvious conclusion would be that it has requirements in both area, but the fact remains that the interplay between thought, practice, and result is still unclear. The dilemma between graphing as a mental covert event as opposed to graphing as an overt physical practice can be extended to include almost aspect of learning. Can a student learn anything without actually doing it? While the philosophical considerations might seem displaced or overly cerebral, most instruction today assumes that knowledge can be delivered, thus gained, by incidental contact with the content. The fact that reading is still the predominant form of learning (and teaching) is a testament to the current norm where it is assumed that learners can glean purposeful information from text. Research into experiential learning and learning as an active process (Hansen, 2000) has highlighted the importance of placing the learner in meaningful contact with the content thus fostering more interaction between learner and content. Unfortunately, even with research to the contrary, the real-world learning environment still relies on instruction via observation. This reliance is probably due to the fact that it is programmatically very difficult to create environments, online in this case, that mimic the personal interaction that offline human environments possess.

To maximize the online potential, environments must be created such that meaningful interaction and feedback are overt and are at the forefront of instruction. This will undoubtedly increase the discourse between the practical and philosophical, and hopefully lessen the distance between the two.

Summary

This chapter reviewed pertinent literature and set the context for the current study. The literature made a strong case for Developmental Research, and studying the development of an environment that pairs direct instruction and simulation should add valuable insight to research literature.

The literature recognizes the need for diversity in the creation of effective instruction. Different strategies and techniques can be used to increase the richness of the instruction. Adaptive instructional techniques and contingent guided practice are only two of the tools that can be used to make instruction effective and meaningful for learners. The techniques reviewed compliment the direct instruction component and the simulation component, resulting in an environment where multiple senses are engaged to facilitate learning.

The model-based systematic approach to instruction was used as a framework for developing LOGIS. The reviewed literature makes a strong case for the use of a systems approach to design. The use of the ADDIE model should provide structure, accountability, and transparency to the development process, increasing the probability that the end result will be both effective and repeatable.

The current study is rooted in both behavioral and cognitive doctrines. Behavioral requirements such as observability are emphasized within both behavioral and cognitive perspectives revealing the importance of those features. Likewise, cognitive principles related to the process of learning and attitudes are also accommodated within the context of this study. The literature admits the differences between the behavioral and cognitive perspectives but reconciles them such that they are both functional aspects of the same instruction. It is clear that both perspectives have advantages that can be exploited to produce optimal instruction, where the resulting instruction has the best features of both worlds.

CHAPTER THREE

PROPOSED DEVELOPMENT

Chapter Map

This chapter fully describes the proposed development of LOGIS using the ADDIE. This chapter presents the initial stages of development and is consistent with the requirements of Developmental Research. The proposed development can be compared to the actual development (chapter 4) to determine the significant elements in the development process, and to highlight important factors in the creation of effective instruction. This chapter coupled with chapter 4 (Actual Development) forms a complete picture of the LOGIS experience from start to finish.

The ADDIE phases and components are based on Gagne et al. (2005) and are followed verbatim from the model summary presented in Gagne et al. (2005, p. 22). The prescriptiveness with which the phases are followed should provide some insulation from critiques concerning the theoretical and practical adaptation of the model. By following the suggested path, the effect of the model as it pertains to the development of LOGIS can be critically and unbiasedly examined. The following map describes the organization of the chapter:

- Proposed development
    - Chapter map

- Research questions
- The ADDIE model
    - Analysis
    - Design
    - Development
    - Implement
    - Evaluate
- Summary

Research Questions

1. How does the use of the ADDIE model influence the creation of the LOGIS instructional application?

To evaluate the ADDIE process, each phase will be subject to the following five questions based on (Gagne et al., 2005, p. 355):

1. Was the rationale for including or omitting this phase clear?
2. Are all required elements of this phase accounted for?
3. Was this phase temporally cost effective to complete?
4. Are the products from this phase observable and measurable?
5. Are the products from this phase consistent with the overall goal of the instructional application?

These questions will be concisely answered at the end of each phase and the results will be reported in chapter four “Actual Development”.

2. Is LOGIS an effective form of instruction?

The effectiveness of LOGIS will not be determined experimentally. This research question will be addressed in chapter four “Actual Development” and also reported in chapter five “Conclusions”. Data from the one-to-one evaluation, the small-group evaluation, and the field trial will be compiled in an effort to determine the effectiveness or value of the instructional application.

The ADDIE Model

This study investigated the model-based creation of instruction and ultimately sought to determine the effectiveness and merit of the instruction. Because of the complexity involved in describing and subsequently using the model, organization became a paramount issue. Gagne et al. (2005) described the ADDIE model as “an organizing framework”. Not only did the model help in the creation of the instruction, but the model also aided in documentation process. The properties of the model thus forced the researcher to examine the organization of all aspects of the study, including documentation, before actually beginning the study.

This chapter presents a detailed description of the phases, sub-components and workflow of the study. This road map will provide an opportunity to maintain contact with both high and low level design requirements. Each phase will be described and each sub-component explained, and these will be used as a guide to create a logical path for the current study. Not all phases and sub-components will be applicable to this study, but where differences occur, a rationale for exclusion will be provided. In cases where modifications are made to the basic structure of the phases or sub-components, there will be ample explanation for the changes.

Gagne et al. (2005) based their description of the model on the development of a course. This study is purposed at developing an instructional application covering one unit that is comprised of several tutorials. Within this chapter, course and unit related references will be demarcated with parentheses followed by the appropriate replacement word or phrase if necessary. This will result in a visible exposition of the amount of changes required to complete this study.

Analysis Phase

Analysis Component 1

(First determine the needs for which instruction is the solution.)

Graphing is an important component of a course at a large southeastern university. The course is entirely online and reading has proven to be an ineffective form of graphing instruction in the course. Previous efforts by the instructor to create graphing instruction resulted in only modest educational gains, thus there is a need for effective graphing instruction.

Learners are required to learn general graphing concepts and how to create simple, cumulative, bar, and multiple baseline graphs. Based on these parameters, carefully designed instruction paired with the opportunity for practice might fulfill the course requirements and at the same time provide learners with a meaningful learning experience.

Analysis Component 2

(Conduct an instructional analysis to determine the target cognitive, affective, and motor skill goals for the [course] unit.)

The instructional analysis will highlight the knowledge, skills, and attitudes (KSA) that learners are expected to acquire. It is important to note that because new instruction will not be created, generating high-level goals will be a step backwards. Instruction for this particular task currently exists, thus the goals and objectives presented in this section are more akin to reorganization rather than creation. Given that the current study was focused on a unit for which goals existed, the goals presented in this section can be viewed as too specific to be goals, but too general to be objectives. These goals are nonetheless very applicable because they state more than “learner will graph”, thus accelerating the study by eliminating some of the intermediate steps between goals formation and the derivation of objectives.

The instructional goals for the unit were classified based on Gagne et al. (2005). The purpose of this classification is to help the instruction sequencing and organization, and also to help provide focus for the assessment. Although each goal can be placed in multiple classifications, they will be placed in the single class that is best suited for that goal. Table 3 is a representation of both the analysis and the classification of the goals.

Table 3

Initial Goals and Tasks Classifications

Goal	Task classification
Knowledge	
After successfully completing this topic, learners will be able to discriminate among common graphing terms and concepts.	Intellectual Skills – Discrimination
After successfully completing this topic, learners will be able to identify the parts of a graph.	Verbal Information
After successfully completing this topic, learners will be able to describe concepts involved in the control and measurement of behavior.	Verbal Information
After successfully completing this topic, learners will be able to state why graphing data is important.	Verbal Information
Skills	
After successfully completing this topic, learners will be able to construct a multiple baseline graph based on a given dataset.	Intellectual Skills – Problem Solving
After successfully completing this topic, learners will be able to construct a cumulative graph based on a given dataset.	Intellectual Skills – Problem Solving

Goal	Task classification
After successfully completing this topic, learners will be able to construct a semi-logarithmic chart.	Intellectual Skills – Problem Solving
After successfully completing this topic, learners will be able to construct a simple bar graph.	Intellectual Skills – Problem Solving
After successfully completing this topic, learners will be able to construct a simple line graph.	Intellectual Skills – Problem Solving
Attitudes	
After successfully completing this topic, learners will choose to graph data when it is the optimal solution.	Attitude

Analysis Component 3

(Determine what skills the entering learners are expected to have, and which will impact learning in the [course] unit.)

This unit will be used in a course where behavioral principles are used as a foundation for coursework, thus learners are expected to have a fundamental grasp of the environment's role in the shaping and maintaining of behavior. The graphing unit is chapter four in the Alberto and Troutman (2006) textbook, and after progressing through chapters one, two, and three, learners are expected to be familiar with PI as an instructional method and the online nature of the course.

Learners in the course are predominantly juniors and seniors and they are expected to have at least a fundamental idea about graphing, but this is not a requirement. It is reasonable to suggest, given the educational level of the students in the course, that they are at least familiar with the rudiments of simple graphing. It is not necessary that they have in-depth knowledge, but a certain assumed proficiency allows the current study to eliminate, for example, instruction on how to draw a straight line.

Learners are not expected to have any particular disposition or motivation towards graphing. The course is open to students of varying backgrounds, intellectual levels, and preparedness. It is expected, however, that students will be sufficiently motivated to perform well on the unit because performance on the unit's subsequent quiz affects the final course grade.

Analysis Component 4

(Analyze the time available and how much might be accomplished in that period of time. Some authors also recommend an analysis of the context and the resources available.)

This instructional unit has several existing parameters. The total number of tutorials must be kept to a minimum, and the total tutorial time must not exceed 2 hours. The 2-hour time limit was determined based on the results and reactions of learners who completed graphing instruction in prior semesters. The instructional unit will eventually be delivered such that learners can work at their own pace, but for the purposes of this study, instructional time must be kept at or around 2 hours.

The course does not cover bar graphs and semi-logarithmic graphs in detail. Based on the lack of emphasis on these types of graphs, constructing bar graphs and semi-logarithmic charts will be removed from the goals. It must be noted that bar graphs and semi-logarithmic charts pose a significant foreseeable programming problem. Under normal circumstances, these graph types would not be removed but based on the lack of course emphasis and in the interest of time and complexity, the current study will not include bar or semi-logarithmic graphs. The goals are thus restated in Table 4.

Table 4

Revised Goals and Task Classifications

Goal	Task classification
Knowledge	
After successfully completing this topic, learners will be able to discriminate among common graphing terms and concepts.	Intellectual Skills – Discrimination
After successfully completing this topic, learners will be able to identify the parts of a graph.	Verbal Information
After successfully completing this topic, learners will be able to describe concepts involved in the control and measurement of behavior.	Verbal Information
After successfully completing this topic, learners will be able to state why graphing data is important.	Verbal Information
Skills	
After successfully completing this topic, learners will be able to construct a multiple baseline graph based on a given dataset.	Intellectual Skills – Problem Solving

Goal	Task classification
After successfully completing this topic, learners will be able to construct a cumulative graph based on a given dataset.	Intellectual Skills – Problem Solving
After successfully completing this topic, learners will be able to construct a simple line graph.	Intellectual Skills – Problem Solving
Attitudes	
After successfully completing this topic, learners will choose to graph data when it is the optimal solution.	Attitude

Design Phase

Design Component 1

(Translate (course) unit goals into overall performance outcomes, and major objectives [for each unit of the course] for the unit.)

The goals outlined in Table 4 are sufficient for this step, because the emphasis is on the instructional unit not the course.

Design Component 2

(Determine the instructional topics [or units] to be covered, and how much time will be spent on each.)

The topics for the tutorials will be based on the instructional goals established in the Analysis phase of this chapter, where each goal can be considered a topic. It is important to remember that because some graphing instruction already exists, the instructional topics generated are in part a reorganizing and re-labeling of current instruction. Each tutorial will be based on one or more topics in the following manner:

- The Control And Measurement Of Behavior
    - After successfully completing this topic, learners will be able to describe concepts involved in the control and measurement of behavior.
- The Importance Of Graphing
    - After successfully completing this topic, learners will be able to state why graphing data is important.
- Basic Graphing
    - After successfully completing this topic, learners will be able to identify the important parts of a graph.
    - After successfully completing this topic, learners will be able to construct a simple line graph.
- Behavioral Graphing Concepts
    - After successfully completing this topic, learners will be able to discriminate among graphing terms and concepts.
- The Cumulative Graph
    - After successfully completing this topic, learners will be able to construct a cumulative graph based on a given dataset.
- The Multiple Baseline Graph
    - After successfully completing this topic, learners will be able to construct a multiple baseline graph based on a given dataset.

The unit composition is already established, but not the time frame for each tutorial. Although there is an overall time limit of 2 hours for completing the unit, the time limit for each tutorial cannot be reliably determined in advanced. It is important to note that the content and subsequent duration of each tutorial will not be based on random estimates. This study will employ a test/revise strategy, the outcome of which will be an optimized set of tutorials. This strategy is detailed in the *Development* section of this chapter.

Design Component 3

(Sequence the [units] tutorials with regard to the [course] unit objectives.)

The previous Design Component section produced six tutorials and they are logically arranged as follows:

1. The Control And Measurement Of Behavior
2. The Importance Of Graphing
3. Basic Graphing
4. Behavioral Graphing Concepts
5. The Cumulative Graph
6. The Multiple Baseline Graph

The *Basic Graphing* tutorial will contain general graphing concepts not related to behavioral graphing. It will be unique in that its delivery will be dependent on the learners' performance on a pre-assessment measure. Based on the principles of Bayesian Model discussed in the literature review section of this study, a simple pre-assessment will be used to determine if the learner needs to complete the *Basic Graphing* tutorial. It is reasonable to assume that a significant number of learners are already versed in basic graphing techniques, thus an extra tutorial based on content that is already mastered would probably be aversive to a majority of the learners.

There is a possibility that some learners will have an inadequate prerequisite skill base. These learners will be accommodated because the *Basic Graphing* tutorial will be a requirement for them. If the *Basic Graphing* tutorial is a requirement for a learner, it will be delivered at the appropriate time, after the *The Importance of Graphing* tutorial.

Design Component 4

(Flesh out the units of instruction, identifying the major objectives to be achieved during each [unit] tutorial.)

Important concepts or rules are presented for each tutorial in the form of key words or short statements. These concepts will be adjusted based on feedback from the content expert and trials of tutorials.

1. The Control And Measurement Of Behavior
    a. Experimentation
    b. Measurement is an important aspect of the science of behavior
    c. Controlling behavior involves the manipulation of environmental variables
    d. It is important to remain in contact with behavior
2. The Importance Of Graphing
    a. Data and graphing
    b. The advantages of visual displays of data
    c. statistical procedures versus visual presentations
    d. Feedback and its importance
    e. Variations in data

    f. Interpretation of data

3. Basic Graphing
    a. Axes
    b. Coordinates
    c. Point
    d. The origin
    e. Scale of axes
    f. Hatch marks
    g. Slope of a line
    h. The title and legend
    i. Graphing data
4. Behavioral Graphing Concepts
    a. Data scale and path
    b. Scale breaks
    c. Dependent and independent variables
    d. Functional relationship
    e. Trends
    f. The importance of time as a unit of measure
5. The Cumulative Graph
    a. The cumulative record
    b. Upward slope of the depth
    c. Difficulty in reading
    d. Rate of responding and its effect on the graph

6. The Multiple Baseline Graph
    a. Graphing multiple datasets
    b. Starting at the origin
    c. Phases and their indications
    d. The indication of special events

Design Component 5

(Define lessons and learning activities for each [unit] tutorial.)

Each tutorial will follow the Programmed Instruction methodology. This means that the content for each tutorial will be a logical linear presentation of frames addressing each objective. The content programming will adhere to the principles highlighted in the review of the literature, and will be accomplished under the supervision of the content expert.

Each appropriate tutorial will have an accompanying practice task where the learner can practice the content presented in the tutorial. Tutorials 1, 2, and 4 will not have any accompanying practice tasks because they primarily deal with abstract concepts or rules. The practice tasks will be delivered using a simulation. Consistent with the reviewed literature, the simulation will be paired with other forms of instruction, in this case PI, in an effort to increase learning and subsequent performance. The fundamental premise is that good instruction paired with the opportunity to practice will produce improved learning and performance.

The design of the simulation is discussed in the Development phase of this chapter, but there are several parameters that must be defined in this subsection. Based on the reviewed literature, the simulation must

1. be consistent and a logical extension of its paired tutorial.
2. resemble the physical graphing environment, thus it must display a graphing grid and have drawing tools.
3. present stimuli requiring the learner to respond, and forward progress must be contingent upon a correct response.
4. provide the learner with sufficient graphing tools such that the learner can edit previous responses.
5. continuously monitor the learners' performance.

The practice tasks for each appropriate topic are as follows:

1. Basic Graphing (construct a simple line graph based on data provided)
2. The Cumulative Graph (construct a cumulative graph based on the data provided)
3. The Multiple Baseline Graph (construct a multiple baseline graph based on data)

Design Component 6

(Develop specifications for assessment of what students have learned.)

Learners will be assessed based on the 8 general goals created in the *Analysis* phase. Verbal information and discrimination (Knowledge) achievement will be assessed using a posttest containing multiple-choice, alternative-choice, and short-answer items. Problem-solving skills (Skills) acquisition will be assessed by requiring the learner to construct paper-based graphs using provided pencil, graph paper, and data. The learners' attitude towards graphing (Attitudes) will be obtained using a survey. Each assessment will be developed and validated in consultation with content and measurement experts.

Development Phase

Development Component 1

(Make decisions regarding the types of learning activities and materials.)

Consistent with the idea that the development of instruction is not a one-time endeavor, accommodations must be made to continually assess and revise the instruction. LOGIS will contain instructional, practice, assessment, and survey components. The lifecycle of the instruction can almost be considered infinite, thus it is important to facilitate the collection and analysis of data such that the instruction can be continually improved.

The basic decision here is that the survey instrument and the pretest and posttest measures will not be separate attachments, but rather these components will be an integral part of the instruction process.

Development Component 2

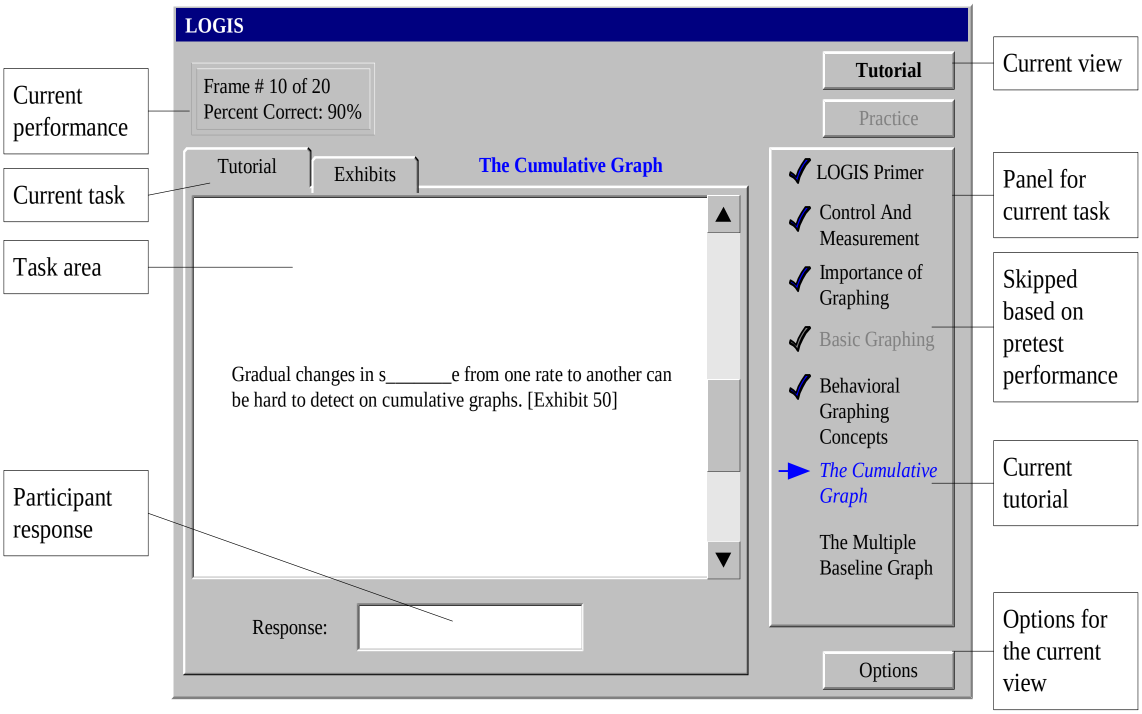
(Prepare draft materials and/or activities.)

The development of the interface and each instrument and measure will be fully described in this section.

The LOGIS interface.

LOGIS will be able to deliver tutorials, practice tasks, tests, and surveys within a single user friendly environment. It will be created as an Applet application using the Java programming language. The Java programming language is powerful, free, highly supported, and browser independent. A backend SQL-server database will be used to house data collected by the application.

Figure 4 is an initial conception of LOGIS and describes some of its proposed components. This is a proposed view of the tutorial screen, where learners will complete tutorials based on behavioral graphing. The application will be storyboarded to optimize the look-and-feel and to ensure the correct interaction of the components.



Frame # 10 of 20
Percent Correct: 90%

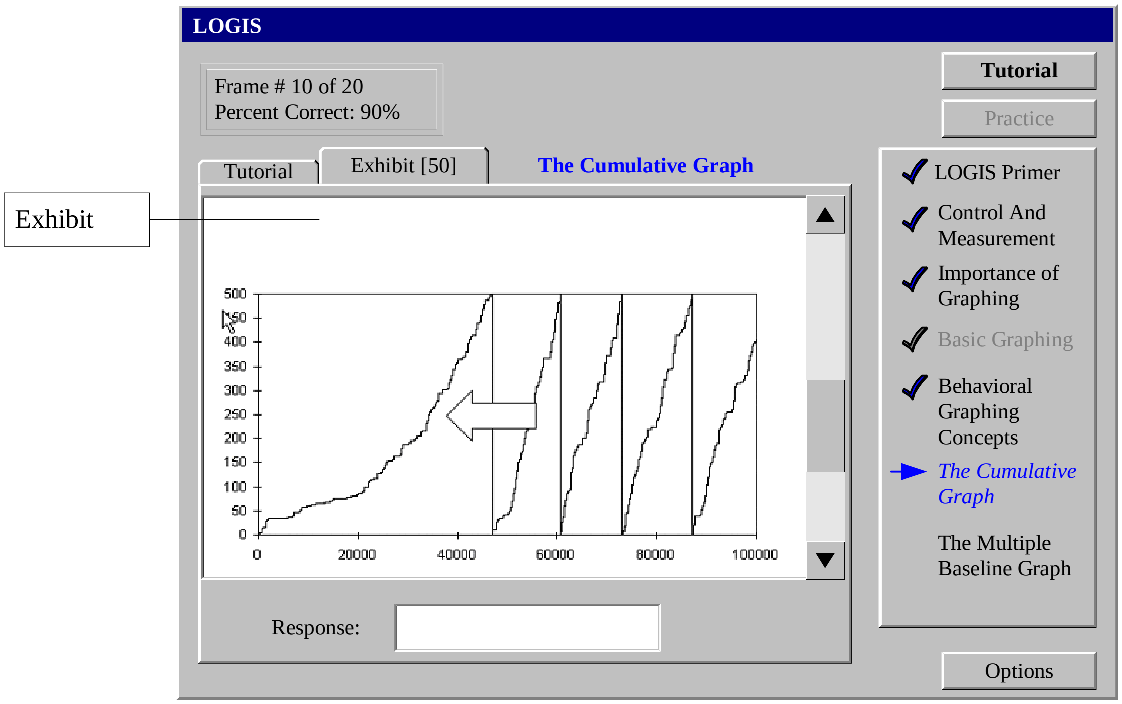
Gradual changes in s______e from one rate to another can be hard to detect on cumulative graphs. [Exhibit 50]

Figure 4. A prototype of the interface for LOGIS showing the tutorial window.

All learners must complete the *LOGIS Primer*. The primer will acquaint the learner with the interface, and explain the function of all available tools. The primer will be refined after the second beta test, based on feedback from the participant interviews.

Learners will complete all available tutorials listed in the *Panel for current task*. If the *Basic Graphing* tutorial label is gray, the learners will be exempt from taking that tutorial. The criterion for exemption from this specific and unique tutorial is perfect performance on the pretest items designed to determine whether or not the learner has already mastered basic graphing concepts. Learners may choose to complete the *Basic Graphing* tutorial, but they are not obligated to do so if they are exempt.

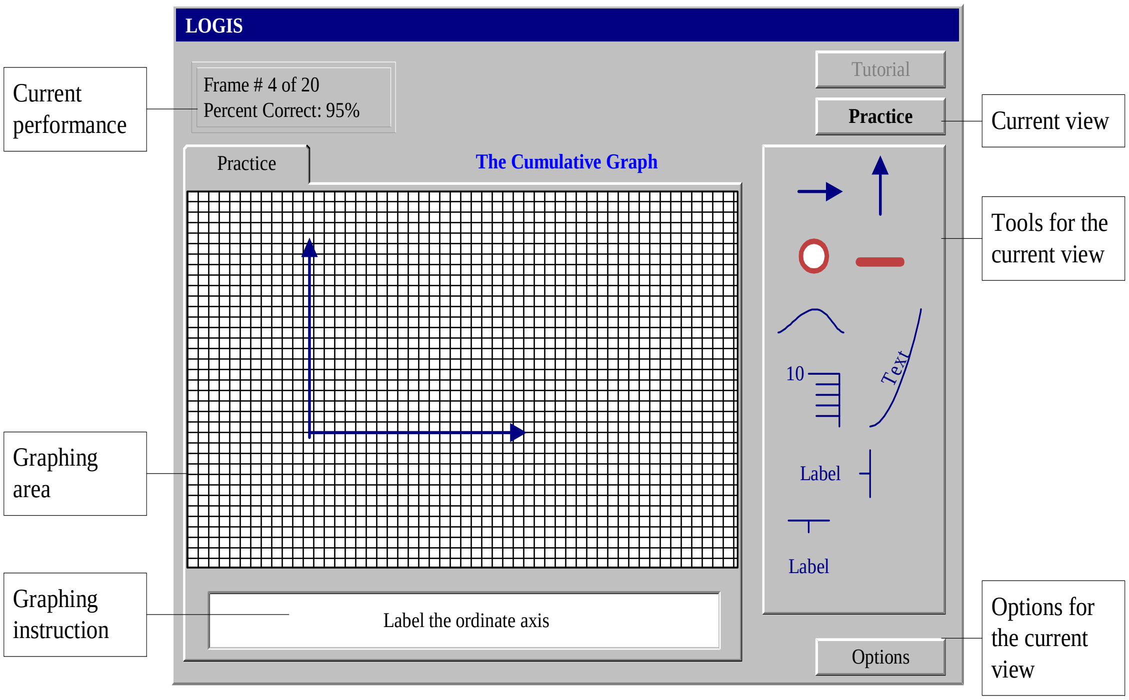
Learners will read the textual content presented in the *Task area* and respond in the designated *Participant response* box. Occasionally, exhibits (see Figure 5) will accompany the frame content. Exhibits are supplemental information that provide examples for increased content clarity. After viewing an exhibit, the learner will be transferred back to the tutorial to continue with the instruction.



Frame # 10 of 20
Percent Correct: 90%

Figure 5. A prototype of the interface for LOGIS showing the exhibit window.

After completing select tutorials, learners will be transferred to the practice screen (see Figure 6) where they will complete a practice task designed to supplement the previously completed tutorial.



Frame # 4 of 20
Percent Correct: 95%

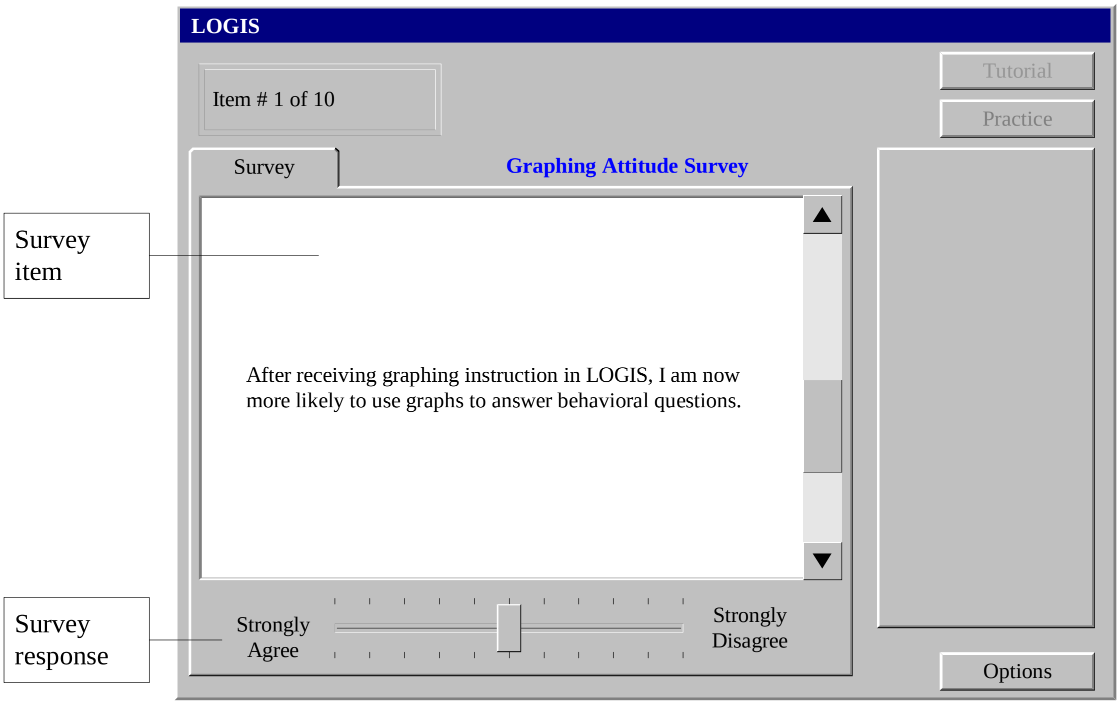
Label the ordinate axis

Figure 6. A prototype of the interface for LOGIS showing the practice window.

During the practice task, learners will perform discrete tasks, and forward stepwise progress will be contingent upon the accurate performance of each step. Practice tasks will involve the step-by-step creation of a specific graph using the tools provided in the *Tools for the current view* panel, where each step must be successfully completed before the practice continues. The directions for each step will be displayed in the *Graphing instruction* box, and learners will complete the step inside the *Graphing area* while receiving feedback on the correctness of their action.

At any point during the tutorial or the practice, learners will be able to see their current progress in the *Current performance* panel. In addition, they will be able to adjust the look-and-feel of the interface using the *Options* button.

The survey component of LOGIS will deliver survey items (see Figure 7) requiring learners' response. The Likert scale which uses a 5-point response scale from strongly agree to strongly disagree, will be replaced by a sliding scale with the same terminal parameters. The slide's position will represent a corresponding number between 1.000 and 5.000. This implementation should increase the precision of the survey instrument because learners will have greater leverage in reporting their opinions.

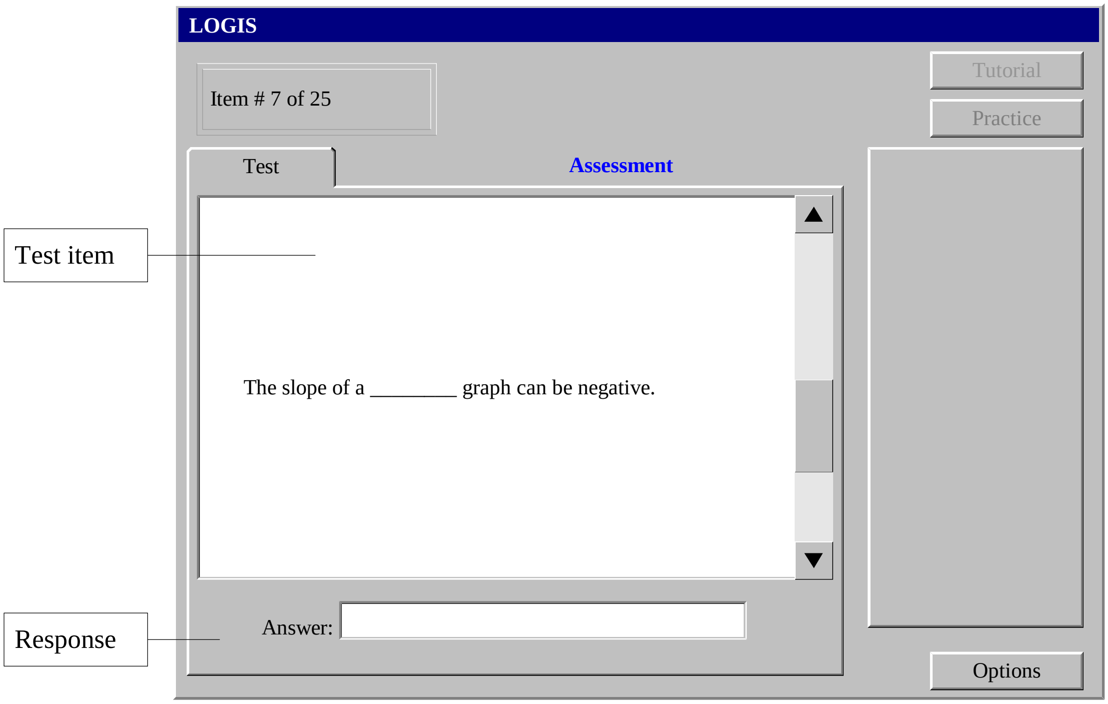


Item # 1 of 10

After receiving graphing instruction in LOGIS, I am now more likely to use graphs to answer behavioral questions.

Figure 7. A prototype of the interface for LOGIS showing the survey window.

The assessment component (see Figure 8) will allow LOGIS to collect pretest and posttest data. Learners will read test items in the *Test item* area then respond in the *Response* box.



The slope of a ________ graph can be negative.

Figure 8. A prototype of the interface for LOGIS showing the knowledge assessment window.

The LOGIS interface is expected to evolve based on expert opinion and advice, and results from the alpha and beta tests.

The LOGIS development process.

LOGIS will be developed using strict guidelines. Firstly, initial content will be created for the tutorials, survey and assessment. These items will be created under the direction of the subject matter and measurement experts. The items will then be alpha tested by the researcher and select individuals who are both knowledgeable about the content and are familiar with the study. The content for the tutorials, survey, and assessment will be revised based on the alpha test results and reactions by the experts and peers.

The content will then be beta tested by students enrolled in one of the behavioral courses. These students will have experienced PI, thus the instructional methodology should be familiar to them. Students will voluntarily complete the content and participation will not affect their course grade in any way. All content items will then be examined and revised, based on feedback from the beta test group.

After the first beta test and subsequent revisions, an application interface will be created. This will be done under the guidance of a subject matter expert and an instructional design expert. Additional survey items will then be created under the guidance of the measurement expert. These new survey items will solicit data on the learners' attitudes towards the interface and overall attitudes towards the application. Next, the practice task will be created under the guidance of the subject matter expert. The revised content, including the new survey items and practice task will then be coupled to the interface and this combination will essentially be LOGIS. LOGIS will then be alpha tested by the researcher and peers and that process will be followed by revisions.

Following the alpha test, LOGIS will be beta tested by students enrolled in one of the behavioral courses. These students will have experienced PI but not LOGIS and they will be different from the first beta testers. Similar to the first beta testers, these students will be under performance based course contingencies.

After completing the beta test, select participants will be randomly chosen and verbally interviewed about LOGIS and subsequently its interface. The interviews will not be taped and the researcher will be the sole interviewer. The interview will be based on three questions:

- What did you like most about (LOGIS/the interface)?
- What did you like least about (LOGIS/the interface)?
- How would you make (LOGIS/the interface) better?

Important points from these interviews will be recorded on paper and used in the revision process.

The development of LOGIS is fully described by Figure 9:

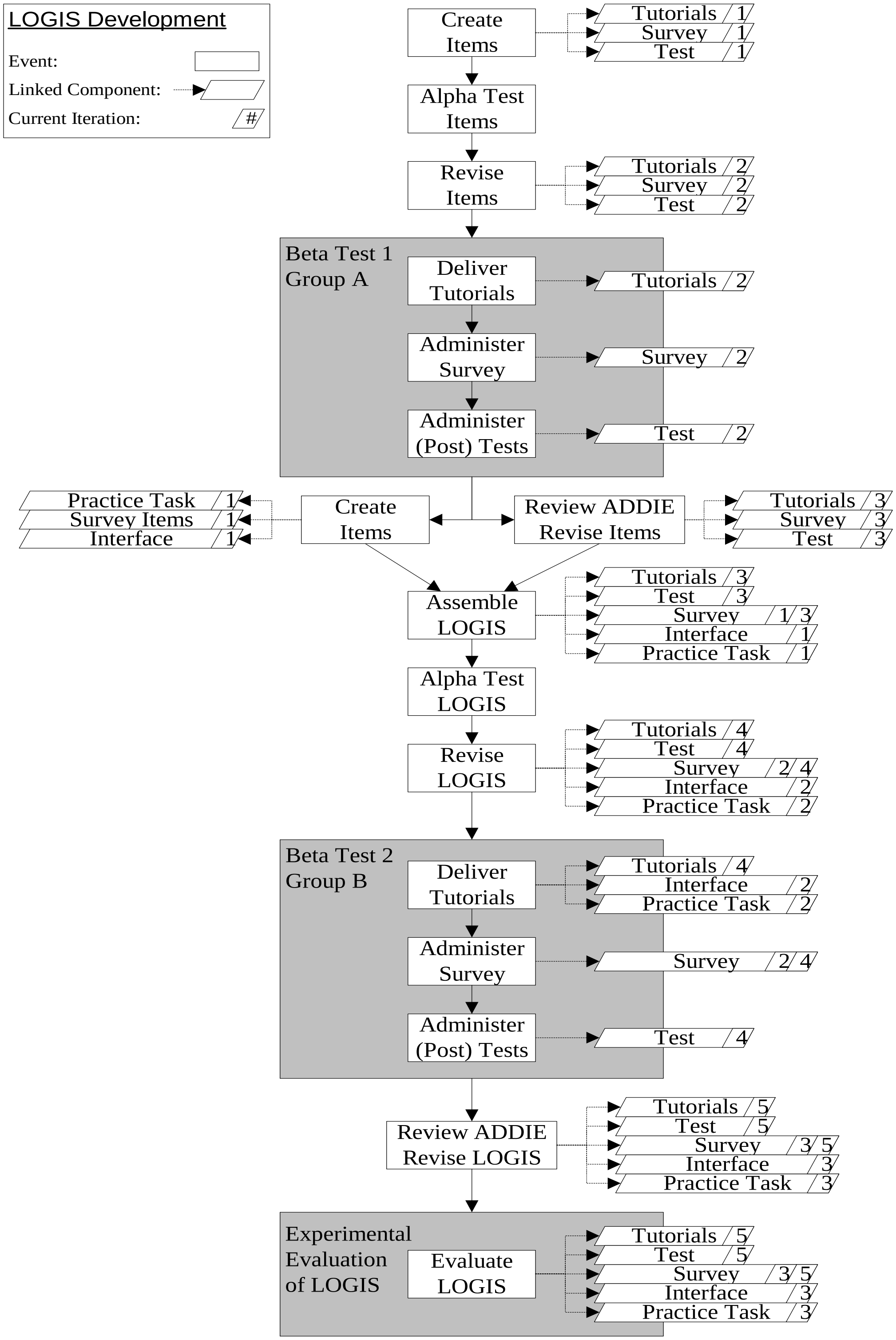


Figure 9. A flowchart of the LOGIS components and instruments development iterations.

The Knowledge Assessment development.

The LOGIS assessment component will focus on two categories, knowledge assessment and skill assessment. The knowledge assessment instrument will be developed based on input from both the subject matter expert and measurement expert.

The knowledge assessment instrument will be developed based on the 10-step process outlined by Crocker and Algina (1986, p. 66):

1. Identify the primary purpose(s) for which the test measurements will be used
2. Identify behaviors that represent the construct or define the domain
3. Prepare a set of test specifications, delineating the proportion of items that should focus on each type of behavior identified in Step 2
4. Construct an initial pool of items
5. Have items reviewed (and revised as necessary)
6. Hold preliminary item tryouts (and revise as necessary)
7. Field test the items on a large sample representative of the examinee population for whom the test is intended
8. Determine statistical properties of the items, and when appropriate, eliminate items that do not meet preestablished criteria
9. Design and conduct reliability and validity studies for the final form of the test
10. Develop guidelines for administration, scoring, and interpretation of the test scores (e.g., prepare norm tables, suggest recommended cutting scores or standards for performance, etc.).

Each step will be completed before the assessment is used in the evaluation. Below is a detailed description of each of the 10 steps:

1. Identify the primary purpose(s) for which the test measurements will be used

This assessment will serve two purposes. Firstly, it will be used to discriminate among learners, using performance as a means of discerning the levels of content mastery that each learner has acquired. Secondly, it will be used as a diagnostic test to determine if a learner has qualified to skip the first tutorial (*Basic Graphing*).

The different objectives, discrimination and diagnostic, are somewhat disjoint and imply different levels of assessment difficulty. This means that ideally the discrimination items will be of medium difficulty where variances in performance will be maximized, and diagnostic items will be somewhat easy such that problem areas are revealed. For the purposes of the current study, diagnostic items will be at least medium difficulty because the objective is to determine if a learner's basic graphing skills are strong enough to warrant skipping the first tutorial.

2. Identify behaviors that represent the construct or define the domain

This assessment will reveal the absolute performance levels of the learners. The behaviors that define this domain are based on the major goals for each tutorial that are defined in the Design phase of this chapter and listed in Appendix A.

Each major objective will have corresponding frames (individual presentations of content), and these frames will be the foundation for the items. This eliminates the need for item specification because the frames themselves contain stimuli that must be mastered during instruction and demonstrated during the assessment. It is important to note that frames will not simply be copied to the assessment. Normally tutorial frames have formal or thematic prompts present, or they may contain references to previous or subsequent frames. This means that the content of each frame will be examined for the potential to be adapted to the assessment. The major advantage of using the frames' content as the foundation for the assessment items is that this method will help keep the assessment items consistent with the instructional content.

It is important to point out that the value of the assessment is not diminished by the initial loose specification items. Crocker and Algina (1986) proposed that as specificity increases, practicality may decrease and that flaws in tightly specified structures may propagate to every item created. They suggested that a certain level of subjectivity and leeway is needed when creating item if issues of practicality and flaw avoidance are to be maximized.

3. Prepare a set of test specifications, delineating the proportion of items that should focus on each type of behavior identified in Step 2

Crocker and Algina (1986) used Bloom's (1956) taxonomy when determining assessment specifications. Bloom is not treated within the current study but Gagne et al. (2005, p. 61) provided a comparison (see Table 5) of Bloom's taxonomy and Gagne's classification scheme:

Table 5

A Comparison of Bloom's Taxonomy and Gagne's Classification Scheme

Bloom	Gagne
Evaluation	Cognitive Strategy, problem solving, rule using
Synthesis	Problem solving
Analysis	Rule using
Application	Rule using

Bloom	Gagne
Comprehension	Defined concepts, concrete concepts, and discriminations
Knowledge	Verbal Information

It is evident that there is overlap between the two classifications, thus, this study will use Gagne’s classification as a base for the assessment specification.

Upon inspection, Gagne et al.’s (2005) classification reveals two strands, Rule-using and Non Rule-using (Gagne considered Verbal Information a separate learning domain). Rule-using is subsumed by problem-solving, thus problem-solving will be the first class. The assessment specifications will be based on these two classes: 1) problem-solving and 2) defined concepts, concrete concepts, and discriminations. This implies that each item on the assessment will be classified as either problem-solving or defined concepts, concrete concepts, and discriminations.

Problem-solving items will compose 70% of the assessment and 30% percent will be defined concepts, concrete concepts, and discriminations items. This distribution describes the importance of higher order activities but does not ignore lower level requirements. The weights of the classes may be adjusted based on expert advice.

The assessment will contain items from each major goal area in each tutorial and will be distributed according to Table 6.

Table 6

Tutorial Weight Distribution

Tutorial	Weight
The Control And Measurement Of Behavior	20%
The Importance Of Graphing	20%
Basic Graphing	0%
Behavioral Graphing Concepts	20%
The Cumulative Graph	20%
The Multiple Baseline Graph	20%

The decision to exclude *Basic Graphing* from the assessment is based on the fact that *Basic Graphing* is a prerequisite for *Behavioral Graphing Concepts*, *The Cumulative Graph*, and *The Multiple Baseline Graph*. Assessing the last three goals sufficiently reveals *Basic Graphing*.

It is expected that a pool of items will be created for each tutorial, covering each major goal. These items will be trial tested during the two beta tests and reduced such that the predetermined distribution for the items and item class is achieved.

4. Construct an initial pool of items

The initial pool of items will be constructed by first considering each frame (stimulus and correct response) of every tutorial and drawing from those frames an appropriate sample representing each major objective. Each item in the sample will then be modified removing cues, prompts, and inappropriate external references. Next, each item will be assigned a class based on the item’s content. It is expected that some statements will be generally more suited for a particular class and they will be assigned accordingly. Each statement will then be examined and an appropriate item format will be assigned. The format will include: alternative-choice, for example true/false and yes/no; multiple-choice; and short-answers. The format assignment will be based on the item's content and class because some formats are more suitable for certain classes, likewise each class has optimal formats.

After the format assignment is complete, each item will be finalized based on the guidelines and checklist developed by Reynolds, Livingston, and Willson (2006, p. 205). The alternate-choice items, for example true/false items, will be developed using the guidelines in Appendix B. After the alternative-choice items have been created, a checklist (Reynolds et al., 2006, p. 207) will be used to finalize those items. If any statement is unchecked, the item will be revised. The checklist is listed in Appendix C.

The multiple-choice items will be developed using the Reynolds et al. (2006, p. 190) guidelines listed in Appendix D. The printed format of the multiple-choice items is very important, this aspect will be guided by Reynolds et al. (2006, p. 190) and will determine the completeness of the first guideline "Use a printed format that makes the item as clear as possible". The print format guidelines are listed in Appendix E. After the multiple-choice items have been created, a checklist (Reynolds et al., 2006, p. 199) will be used to finalize the items. If any statement is unchecked, the item will be revised. The checklist is listed in Appendix F.

The short-answer items will be developed using the Reynolds et al. (2006, p. 232) guidelines listed in Appendix G. After the short-answer items have been created, a checklist (Reynolds et al., 2006, p. 233) will be used to finalize the items. If any statement is unchecked, the item will be revised. The checklist is listed in Appendix H.

These guidelines will form a basis for the development of the various items. Where appropriate, the guidelines will be modified or adapted to suit the requirements of the current study.

5. Have items reviewed (and revised as necessary)

The pool of items will be reviewed by experts. The review will be based in part on the following considerations:

- Accuracy
- Appropriateness or relevance to test specifications
- Technical item-construction flaws
- Grammar
- Offensiveness or appearance of "bias"
- Level of readability

(Crocker & Algina, 1986, p. 81)

6. Hold preliminary item tryouts (and revise as necessary)

Preliminary tryouts will be completed in first beta test, and the item pool is expected to be significantly larger than the target amount. Statistical procedures and expert examination will reduce the number of items in the pool before the second beta test.

7. Field test the items on a large sample representative of the examinee population for whom the test is intended

This field test will be completed in the second beta test. The number of items in the pool will be significantly less than in the first beta test. After the second beta test, statistical procedures and expert examination will reduce the number of items in the pool to the target number of items for this assessment.

8. Determine statistical properties of the items, and when appropriate, eliminate items that do not meet preestablished criteria

The Item Difficulty Index will be calculated for each item in the pool and they will each be judged based on optimal *p* values provided by Reynolds et al. (2006, p. 144), where the *p* value is the proportion of learners that answered the item correctly. The desired *p* value for alternate-choice items is .85, for short-answers it is .50, and for multiple-choice items it is .74 for items with four distracters and .69 for items with five distracters. Items that are either too high or too low will be examined and possibly modified or removed from the pool.

Cronbach's Alpha will be also be used to estimate the internal consistency of the instrument. Items with low item-total correlations will be modified or removed.

Multiple-choice items will be subject to Distracter Analysis to ensure that the distracters are performing well. Distracters are supposed to attract learners in the bottom 50% group while correct answers are supposed to attract the top 50% group. Distracter Analysis will entail a manual examination of the data and decision regarding revision or elimination will be made with expert consultation.

The beta tests are expected to produce data which will be examined using a table similar to Figure 10 allows for the immediate examination of all data points, facilitating the decision-making process. Multiple-choice items totals will be housed in fields 1 through 5 depending on the number of distracters in the item. Alternate-choice items totals will be housed in field 1 and field 2. Short-answers items responses will not be entered into the table.

Item	Item Responses					item-total correlation	p
	1	2	3	4	5		

Figure 10. A prototype knowledge assessment data worksheet.

9. Design and conduct reliability and validity studies for the final form of the test

Validity and reliability studies will be conducted at both beta tests. Studies conducted after the first beta test are not the norm, but in this case it is necessary to assess the progress of the instrument at every step because there are only two opportunities for beta testing.

Validity refers to the correctness or accuracy of interpretations made based on collected data (Reynolds et al., 2006). The validation process, which is the collection and analysis of evidence to support the interpretation, is divided into three groups: Content-related, Construct-related and Criterion-related. Construct-related evidence of validity can be viewed as a superset of both content-related evidence, and where applicable, criterion-related evidence (Fraenkel & Wallen, 2006; Reynolds et al., 2006; Trochim, 2002). The investigation of the construct usually involves an in-depth analysis of the content, its purpose, and all pertinent criteria. To avoid confusion among the different terms, this study will consider content-related evidence as sufficient for the validation process.

Content-related evidence of validity for the assessment will be determined prior to each beta test. It is important to note that the beta tests involve a pool of items and although the pool is not the target assessment, it is important that the validation process be applied at each step to ensure an optimal final product.

Content and measurement experts will be asked to examine items and determine if the items address the content domain. The experts will be presented with the item, purpose, major goal, and class. They will be asked to determine if the item mirrors its purpose, addresses its major goal, and if it is both appropriate and a good representation of the class to which it is assigned. The items will then be revised based on feedback from the experts.

Reliability refers to the consistency of the obtained data (Fraenkel & Wallen, 2006). A reliability coefficient estimate will be calculated after each beta test. An Internal Consistency method will be used because only one administration of the assessment measure will occur. The knowledge assessment will use the alpha coefficient (Cronbach's Alpha).

10. Develop guidelines for administration, scoring, and interpretation of the test scores (e.g., prepare norm tables, suggest recommended cutting scores or standards for performance, etc.)

The knowledge assessment will be delivered online one item at a time without the opportunity to backtrack. This will reduce the use of clues from previous items and provide a more controlled testing environment. The assessment will be electronically scored and learners will not have running tally of their score when they are taking the assessment.

It is expected that participants in the current study will make mistakes on the short-answer items. Errors including incorrect spelling, wrong tense, and singular versus plural, are just a few of the mistakes that are expected. Each response for every participant will be examined prior to any data analysis. If obvious errors are present, credit will be given to the participant and their score will be updated

The Skills Assessment development.

The skills assessment instrument will be developed based on a modified version of the ten step process outlined by Crocker and Algina (1986, p. 66):

1. Identify the primary purpose(s) for which the test measurements will be used

This assessment will be used to discriminate among learners, using performance as a means of discerning the levels of skill mastery that each learner has acquired.

2. Identify behaviors that represent the construct or define the domain

LOGIS delivers instruction to learners such that the terminal skill behavior is the construction of cumulative graphs and multiple baseline graphs. These two behaviors form the domain for the skills assessment.

3. Prepare a set of test specifications, delineating the proportion of items that should focus on each type of behavior identified in Step 2

The skills assessment will have four required items, two cumulative graphs of similar difficulty and two multiple baseline graphs of similar difficulty, where the difficult of the items will be assessed by the content expert. These graphs will represent an even distribution of the domain behaviors.

The practice tasks within LOGIS are the primary skills instruction agents. They will contain discrete steps that lead to the creation of a correct graph. Each step is in essence an item and can be treated as such. These items (steps) will be collected for both the cumulative graph tasks and the multiple-baseline tasks and they will be used as the basis for the item specifications for the skills assessment. This means that all items in the LOGIS practice tasks will be examined and possibly adapted to become assessment points in the skills assessment.

4. Have items reviewed (and revised as necessary)

The four skills items will be reviewed by experts. The review will be based on Crocker and Algina (1986, p. 81):

- Accuracy
- Appropriateness or relevance to test specifications
- Technical item-construction flaws
- Grammar
- Offensiveness or appearance of "bias"
- Level of readability

After the skills items have been created, the checklist listed in Appendix I will be used to finalize the items. If any statement is unchecked, the items will be revised.

5. Hold preliminary item tryouts (and revise as necessary)

Preliminary tryouts will be completed in the first beta test and the four skills items will be adjusted based on performance. The revision will take the target performance criterion for the instructional unit into consideration. It is important to note that the terminal skill performance requirement is that the learners construct correct graphs. To this end, revisions will only occur if they do not jeopardize the overall goals and target mastery levels of the instruction.

6. Field test the items on a large sample representative of the examinee population for whom the test is intended

Field testing will occur in the second beta test, followed by further revision if necessary.

7. Design and conduct reliability and validity studies for the final form of the test

Validity and reliability studies will follow the same theme as those used in the knowledge assessment. Content-related evidence of validity for the assessment will be determined prior to each beta test. Content and measurement experts will be asked to examine the four skills items and determine if they address the content domain, that is, if they require participants to demonstrate what they have learned, and if they are of the appropriate skill level. The skills items will be revised based on feedback from the experts.

Three raters will independently score each skill assessment item. The purpose of multiple raters is to ensure that the grading is consistent and without bias. Two scores will be examined to determine the reliability of the skills assessment, and they will represent consensus and consistency estimates. It is important to return two scores because inter-rater agreement (consensus) does not necessarily imply inter-rater consistency (consistency) and vice-versa (Stemler, 2004). Both estimates will be used in tandem because the goal is to produce scores with high agreement and consistency, justifying the averaging of raters score to produce an aggregate score for each participant.

First, the inter-rater agreement will be calculated. This percent score will reflect the degree of agreement among the three raters. Secondly, an Intraclass Correlation Coefficient (ICC) will be calculated using the two-way random effects model. This model was chosen based on the flowchart for selecting the appropriate ICC (McGraw & Wong, 1996, p. 40). The raters in this study will be viewed as both a second factor and as a random sample. Their variability will be considered relevant and they will be selected from a pool of available raters. The result is a two-way random effects model, and although it is computationally similar to the mixed effects method where raters are considered fixed and their variability irrelevant, the two-way random effects model is generalizable to a larger rater population. If the ICC falls below .80 or the agreement falls below .90 (Fraenkel & Wallen, 2006), the rubric will be examined and possibly modified, the assessment will re-graded, and if necessary the raters will be retrained. The exact cutoff limits values for the agreement score and the ICC estimate will be carefully considered and possibly reevaluated after data has been collected. It must be noted that it is entirely possible for the assessment to be reliable even if .90 agreement and .80 correlations does not occur.

The three raters will be selected and trained prior to the first beta test. The first two raters will be knowledgeable about the content and familiar with the PI instructional method. The third rater will be from outside the content field and will not be familiar with either the content or PI as an instructional method. This representation of raters does not violate the randomness criterion for the ICC model because these raters are simply the individuals chosen.

8. Develop guidelines for administration, scoring, and interpretation of the test scores (e.g., prepare norm tables, suggest recommended cutting scores or standards for performance, etc.)

The skills assessment will require that participants manually create four graphs. These graphs, two cumulative and two multiple-baseline graphs, will be scored by three raters independently. The graphs will be scored based on a rubric that will be developed based on the item specifications for the assessment. The rater will assign points (0, 1, 2, or 3) to each item on the rubric, for every participant. The point values are described in Table 7.

Table 7

Points Allocation for Items on the Skills Assessment Rubric

Points	Meaning
0	The item's criterion is completely absent
1	The item's criterion is minimally fulfilled. The item's criterion is significantly over-represented or under-represented
2	The item's criterion is satisfactorily represented. The item's criterion is minimally over-represented or minimally under-represented
3	The item's criterion is perfectly represented

The points earned by a participant will be summed and a percent score will be calculated based on the participant's earned points and the total possible points for the graph. The maximum number of points will occur if the participant is awarded three points for every item on the rubric. The percent scores for each of the four graphs will be averaged and the resulting score will be participant's overall score on the skills assessment.

The Survey development.

The Graphing Attitude Survey instrument will be developed based on a modified version of the ten step process outlined by Crocker and Algina (1986, p. 66):

1. Identify the primary purpose(s) for which the test measurements will be used

The Graphing Attitude Survey will be used to describe the attitudinal characteristics of the participants.

2. Identify behaviors that represent the construct or define the domain

The survey items will be categorized into three groups, attitude towards graphing, attitude towards the LOGIS interface, and attitude towards the LOGIS application. Attitude towards graphing will reflect the learners' willingness to use graphing techniques to solve problems, where choosing to use graphing techniques is equated to a positive attitude towards graphing. Consistent with the review of the literature, the attitude construct will be operationalized as *choice*.

Items involving attitude towards the LOGIS interface will not be created until after the first beta test because the LOGIS will not be assembled until that time. Attitude towards the interface will be measured by learners' report on various aspects of the interface; choosing to use the interface is equated to a positive attitude towards the interface.

The learners' attitude towards the LOGIS application is a measure of the learners' perception of the value or worth of LOGIS. Those who perceive LOGIS to be a useful and effective application will more than likely respond positively to items in this category, choosing LOGIS if given the choice of instructional applications.

3. Prepare a set of test specifications, delineating the proportion of items that should focus on each type of behavior identified in Step 2

Each item in the survey will belong to only one category and the categories will have an equal number of items.

4. Construct an initial pool of items

The initial pool of items will focus on the participants' attitude towards graphing. This pool will be constructed based on the guidelines provided by Crocker and Algina (1986, p. 80). The guidelines are listed in Appendix J. After the survey items have been created, the checklist listed in Appendix K will be used to finalize the items. If any statement is unchecked, the items will be revised.

5. Have items reviewed (and revised as necessary)

The survey items will be reviewed by experts. The review will be based on Crocker and Algina (1986, p. 81):

- Accuracy
- Appropriateness or relevance to test specifications
- Technical item-construction flaws
- Grammar
- Offensiveness or appearance of "bias"
- Level of readability

6. Hold preliminary item tryouts (and revise as necessary)

Preliminary tryouts will be completed in the first beta test, and the item pool is expected to be significantly larger than the target amount. Expert examination of the resulting descriptive statistics will reduce the number of items in the pool before the second beta test.

7. Field test the items on a large sample representative of the examinee population for whom the test is intended

This field test will be completed in the second beta test. The number of items in the pool will be significantly less than in the first beta test. At this point, survey items focused on attitude towards LOGIS and attitude towards the LOGIS interface will be added to the pool. After the second beta test, expert examination of the resulting descriptive statistics will reduce the number of items in the pool to the target number of items for this survey.

8. Determine statistical properties of the items, and when appropriate, eliminate items that do not meet pre-established criteria

The survey data will be statistically examined using exploratory factor analysis where items are associated a priori with factors. To facilitate interpretation, the output will be rotated using the Varimax rotation method.

9. Design and conduct reliability and validity studies for the final form of the test

Validity and reliability studies will be conducted after both beta tests and will follow a theme similar to the knowledge assessment. The theoretical background for the procedures is discussed in the The Knowledge Assessment development section.

Prior to each beta test, content and measurement experts will be asked to examine items and determine if they address the content domain. The experts will be presented with the item and the category. They will be asked to determine if the item is both appropriate and a good representation of the category to which it is assigned. The items will be revised based on feedback from the experts.

A reliability coefficient estimate will be calculated after each beta test. An Internal Consistency method will be used because only one administration of the assessment measure will occur. The survey will use the alpha coefficient (Cronbach's Alpha) to determine reliability and aid in the revision process.

10. Develop guidelines for administration, scoring, and interpretation of the test scores (e.g., prepare norm tables, suggest recommended cutting scores or standards for performance, etc.)

The Graphing Attitude Survey will be delivered electronically. A digital slide based on a 1.000 to 5.000 scale with three decimal places accuracy will be used in lieu of the Likert scale. The scale will range from strongly positive to strongly negative with a 3.000 being neutral. Similar to the Likert scale, the slide is a bipolar, unidimensional, interval measurement scale that can be used for structured response formats (Trochim, 2002). The only item that will not use this method is the survey item that requests general feedback on the interface. That item will be an unstructured response format; therefore it will only be used as a guide in revising the interface. Using a digital slide instead of the usual five point Likert Scale will increase the accuracy of the responses without compromising the goals of the survey.

The results from the Graphing Attitude Survey will be a part of the estimation of the usefulness and value of LOGIS. To facilitate this estimation, an aggregate score will be calculated for the attitude towards graphing category. The aggregate score will be an average of the scores of all the items in the attitude towards graphing category. This means that along with the knowledge and skills assessments scores for the pre and posttests, each participant will have one score representing graphing attitude before the treatment (pre-survey) and one score representing graphing attitude after the treatment (post-survey). The pre and post surveys will contain the same items for the attitude towards graphing category, but only the post survey will have items from the attitude towards the LOGIS interface and attitude towards LOGIS categories.

The graphing attitude aggregate score is necessary for the estimation of the value of LOGIS and it cannot include either the attitude towards the LOGIS interface score or the attitude towards LOGIS score because neither can be assessed in the pre-survey.

Threats to assessment validity

Trochim (2002) listed several threats to validity and they must be account for within the current study. The current study contains two constructs that must be explicitly defined and operationalized. A part of this study is the estimation of the value of the instructional treatment and its relationship to learning. The two constructs, instruction and learning have been theoretically examined and translated into LOGIS. LOGIS is an application (instructional construct) that produces an observable outcome (learning construct). The fundamental constructs of instruction and learning form the basis for threat analysis of the assessments:

- Inadequate Preoperational Explication of Constructs
    - This threat is a result of poor initial definitions and operationalization of the constructs. It is expected that expert critique will reveal inadequate definitions of constructs if they exist.
- Mono-Operation Bias

  - This threat does not pertain to the assessment measures. Although the assessment measures are an integral part of LOGIS, it is accepted that different versions of LOGIS might yield different performance outcomes. The step-by-step creation of LOGIS reduces this threat because multiple opportunities exist to determine if LOGIS is performing well.
- Mono-Method Bias
  - Multiple opportunities to analyze the observations exist thereby reducing this threat.
- Interaction of Different Treatments
  - This is not expected to be a significant threat, because the treatment is unique at least within the participants' course.
- Interaction of Testing and Treatment
  - This is a valid and accepted threat because the pretest assessment instrument might bias the treatment and subsequent performance outcome. The delay between the pretest and the treatment is not sufficient to reduce this threat.
- Restricted Generalizability Across Constructs
  - This threat is accepted because the current study does not readily identify all possible constructs that may be impacted by the treatment.
- Confounding Constructs and Levels of Constructs
  - This threat is accepted because the current study does not readily identify all possible constructs that may be impacted by the treatment, nor all possible forms of the treatment. LOGIS is not a solution to any construct

other than those described by the current study.

The “Social” threats to construct validity will also be addressed. Hypothesis Guessing, Evaluation Apprehension, and Experimenter Expectancies are not expected to be significant threats. Every attempt will be made to make the assessments as unobtrusive as possible and make their delivery as uniform and consistent as possible.

Threats to the survey validity

The survey is based primarily on the attitude construct, which has been defined and operationalized. In addition to the threats identified for the assessments, the survey is susceptible to several validity threats. Mortality, Location, Instrumentation and Instrument decay (Fraenkel & Wallen, 2006) are all pertinent to surveys but they are not expected to be significant for this survey. Mortality is a concern for longitudinal studies where missing participants translate into missing data. Location is not expected to be a factor because all participants will complete the survey in the familiar and non-threatening computer laboratory. Instrumentation concerns are expected to be minimized because of the multiple refinements that will occur. Instrument decay is not expected to be a factor because participants will have as much time as they need to complete the survey.

Development Component 3

(Try out materials and activities with target audience members.)

This sub-component is treated in the previous sub-component “Prepare draft materials and/or activities”.

Development Component 4

(Revise, refine, and produce materials and activities.)

This sub-component is treated in the previous sub-component “Prepare draft materials and/or activities”.

Development Component 5

(Produce instructor training or adjunct materials.)

This sub-component is beyond the scope of the current study.

Implement Phase

Implement Component 1

(Market materials for adoption by teachers or students.)

This sub-component is not applicable to the current study.

Implement Component 2

(Provide help or support as needed.)

This sub-component is not applicable to the current study.

Evaluate Phase

Evaluate Component 1

(Implement plans for student evaluation.)

Learner evaluation is an integral part of LOGIS. The knowledge, skills and attitude measures will be used in part to determine the effectiveness and value of LOGIS. The second research question “Is LOGIS an effective form of instruction?” will be determined after the field trial is completed and will be based on non-experimental analysis. The field trial procedures are fully discussed in the section “Implement plans for program evaluation.”

Effectiveness will be determined based on the (KSA) goals defined earlier in this Chapter. Those goals will serve as the basis for quantitative judgment on the effectiveness of the instruction, in essence whether or not learning occurred. The goals will be measured by a posttest for the knowledge component (K), a posttest for the skills component (S), and a survey (Graphing Attitude Survey) for the attitude component (A).

Effectiveness will be based on academic parameters but statistical data will be used to support estimates of usefulness or value. This means that LOGIS will be considered useful and effective if it produces educationally significant differences from pretest to posttest. Educationally significant differences will be characterized by at minimum a mean 10% increase in performance from the pretest to the posttest. The 10% mark is not arbitrary, an increase of 10% is a guarantee that a student's grade will increase by one letter grade, thus an educationally significant increase. The attitude measure's contribution to effectiveness will also be based on the 10% figure. In this case, the 10% will simply be used for consistency. It must be noted that attitudinal measure of effectiveness will be based solely on the aggregate score of the *attitude towards graphing* category, as described in the "The Survey development" section in this chapter.

Descriptive statistics will be reported in addition to the differences in pretest and posttest scores. The mean and distribution of the scores will add insight into the effectiveness on LOGIS. It must be restated that the intent is not claim that LOGIS *causes* increased performance. The objective is simply to determine an estimate of the application's value or usefulness.

Evaluate Component 2

(Implement plans for program evaluation.)

The formative evaluation of LOGIS will entail a one-to-one evaluation, a small-group evaluation, and a field trial (Dick et al., 2005; Morrison et al., 2004; P. L. Smith & Tillman, 2005). These three formative evaluation steps will be conducted sequentially where the results from one step will be used as a guide to initiate the next step.

The researcher will be responsible for recording key observations and reactions, allowing the participant to focus entirely on the instruction. This will reduce the participant's workload and encourage meaningful interaction between the researcher and the participant.

The revision process will occur after each evaluation. In the case of the one-to-one evaluation, revisions will occur after each participant has completed the instruction. Performance data, survey results, and descriptive evaluation information will be summarized then analyzed and recommended changes will be made. It is important to note that errors, for example typographical errors or broken links, will be corrected immediately. Other errors, for example complex words or phrases, might require adding or removing qualifiers or explanatory sections. If errors appear to result from participant specific attributes, those errors will be noted and the decision to address those errors will be made after all evaluations are complete. The decision to revise participant specific errors will be made based on the researcher's and expert insight, error complexity, available time, and available resources.

The participants for the evaluation phase are expected to come from an undergraduate course at a major southeastern university. The course is expected to have an enrollment of between 80 and 100 students. These students generally have diverse majors, are at different academic levels on the continuum from freshman to senior, and represent different age groups. These students do not have a graphing component in their course, but they will have been exposed to behavioral principles and will be comfortable with PI as an instructional format. It is important to note that no student will take more than one type of evaluation.

The evaluation will be completed in a computer laboratory at a major southeastern university. The computer laboratory contains twenty personal computers arranged in five rows with four machines in each row. There are two types of computers in the laboratory, 13 computers carry the Intel Pentium 350 Megahertz processor and 7 carry the AMD 1.4 Gigahertz processor. The AMD computers are much faster than the Intel computers, but they all run the Windows XP operating system and have similar software packages installed. All computers in the laboratory are connected to the internet via 100 Megabits per second Ethernet connections.

One-to-One Evaluation

The one-to-one evaluation is used to detect obvious errors in instruction and also to obtain initial performance data and user reactions. During this evaluation, the designer interacts with individual users as the users complete the instruction. This interaction provides the avenue through which data and reactions can be observed and recorded (Dick et al., 2005).

Based on recommendations by Dick et al. (2005), three students will be selected to perform this evaluation. These students will represent the upper, middle and lower performers in the course. To select the three students for the evaluation, all students will be ranked based on their current course grade. The first selection will be the student at the middle position of the upper 25% of the rank order. The second selection will be the student at the overall middle position of the rank order. The final selection will be the student at the middle position of the lower 25% of the rank order.

The data collection procedure for the one-to-one evaluation will be based on Dick et al. (2005, p. 283). Descriptive information will be recorded by the researcher and will be gathered using the following questions as the basis for verbal interaction:

1. Did you have any problems with specific words or phrases?
2. Did you have any problems with specific sentences?
3. Did you understand the themes presented in the instruction?
4. How effective was the sequencing of the instruction?
5. How would you describe the delivery pace of the instruction?
6. Was sufficient content presented in each section?
7. Was there sufficient time to complete the instruction?
8. What is your overall reaction to the instruction?
9. What are specific strengths of the instruction?
10. What are specific weaknesses of the instruction?

These questions will encourage the participant to verbalize both strengths and weaknesses of the instruction, as well as provide an opportunity for general comments related to the instructional application.

The researcher will establish rapport with the participant by encouraging the participant to react freely and honestly to the instruction. Interaction between the researcher and the participant is critical to this evaluation. The participant will be asked to verbalize each step, thus facilitating rich and meaningful dialog between the researcher and the participant.

The performance data, survey results, and one-to-one evaluation information for the participant will be combined to form a general picture of the instruction. These results will be used to further refine the application. Based on expected changes in the instruction and differences between participants, observations and results are expected to differ among the one-to-one evaluations.

Small-Group Evaluation

The small-group evaluation is used to test the effectiveness of the one-to-one evaluation, and also to determine if participants can complete the instruction without intervention (Dick et al., 2005).

Based on recommendations by Dick et al. (2005), 9 students will be selected to perform this evaluation. These students will represent the upper, middle and lower performers in the course. To select the nine students for the evaluation, all students will be ranked based on their current course grade. The first selection will be the 3 students at the middle position of the upper 25% of the rank order. The second selection will be the three students at the overall middle position of the rank order. The final selection will be the three students at the middle position of the lower 25% of the rank order.

Unlike the one-to-one evaluation, the researcher will not interact with the participant during the instruction. Except in extreme cases where, for example equipment failure occurs, the researcher will limit interaction until the instruction is complete.

The data collection procedure for the small-group evaluation will be primarily based on the posttest, the survey, and descriptive information (Dick et al., 2005, p. 288). Descriptive information will be recorded by the researcher and will be gathered using the following questions as the basis for verbal interaction:

1. Did you have any problems with specific words or phrases?
2. Did you have any problems with specific sentences?
3. Did you understand the themes presented in the instruction?
4. How effective was the sequencing of the instruction?
5. How would you describe the delivery pace of the instruction?
6. Was sufficient content presented in each section?
7. Was there sufficient time to complete the instruction?
8. What is your overall reaction to the instruction?
9. What are specific strengths of the instruction?
10. What are specific weaknesses of the instruction?

These questions will encourage the participant to verbalize both strengths and weaknesses of the instruction, as well as provide an opportunity for general comments related to the instructional application. The researcher will establish rapport with the participant by encouraging the participant to react freely and honestly to the instruction. This will facilitate meaningful dialog between the researcher and the participant.

The performance data, survey results, and descriptive information provided by the participants will be combined to form a general picture of the instruction. These results will be used to further refine the application.

The participants for the small-group evaluation are not expected to complete the instruction at the same time. This means that minor revisions might occur between evaluations thus observations and results are expected to slightly differ among the evaluations.

Field Trial

The field trial evaluation is used to test the effectiveness of the small-group evaluation, and also to determine if the instruction is effective under normal learning circumstances (Dick et al., 2005).

Based on recommendations by Dick et al. (2005), 20 students will be randomly selected to perform this evaluation. Unlike the one-to-one and small-group evaluations, the researcher will not interact with the participants during the instruction or after the instruction. Except in extreme cases where, for example equipment failure occurs, there will be no interaction with the participant.

The data collection for the field trial evaluation will include the posttest and the survey results (Dick et al., 2005, p. 291). The application will be revised based on the resulting data.

Evaluate Component 3

(Implement plans for unit [course] maintenance and revision.)

This sub-component is not applicable to the current study.

Summary

This chapter outlined the proposed model-based development of LOGIS. Consistent with the goals of Developmental Research, the combination of this chapter (Proposed Development) and chapter four (Actual Development) should present a complete picture of the process that was used to develop LOGIS. The prescriptive nature of the documentation of phases should provide an avenue to answer the research question while helping to organize the development process. It must be noted, that under normal circumstances, the Evaluation phase might not be a summative evaluation. Considering the view that the creation of instruction is a dynamic and ongoing process, the relevance of summative evaluation is questionable. Creation as a process that requires constant evaluation and revision implies that the summative evaluation is simply an in-depth formative evaluation.

CHAPTER FOUR

ACTUAL DEVELOPMENT

Chapter Map

This chapter describes the changes from the proposed development to the actual development of LOGIS. These changes are discussed within the framework of the research questions. Collectively, Chapter Three (Proposed Development) and Chapter Four (Actual Development) present a complete picture of the development process from conception to implementation. The following map describes the organization of this chapter:

- Actual Development
    - Chapter map
    - Research question
    - Analysis Phase Report and Reflection
        - Analysis Phase Critique
    - Design Phase Report and Reflection
        - Design Phase Critique
    - Development Phase Report and Reflection
        - Development Phase Critique
    - Implement Phase Report and Reflection

        - Implement Phase Critique
    - Evaluate Phase Report and Reflection
        - Evaluate Phase Critique
    - Summary

The researcher kept a written journal of observations pertaining to the use of the model, instrument creation, and software development. The purpose of this log was to document the development process and provide a point of reference for future commentary on the development of the application. Using a log was not an initial consideration in the proposed development, but the decision to create an observation log was made at the beginning of the Analysis phase. The log was used to record procedures, observations, thoughts, questions, and comments. A significant portion of the critique of each phase is the result of reflecting on the material written in the log book.

LOGIS took approximately 33 weeks to complete. There were periods of high activity and productivity and there were times when less work was accomplished. This situation made it difficult to determine the exact amount of time that was spent on each phase. General estimates of the time spent on each phase were determined based on the computer-logged dates of files used in the study, logbook entries, and researcher-generated estimates. These estimates are not exact but they provide a basis for commentary on the temporal cost of each phase.

Research Questions

1. How does the use of the ADDIE model influence the creation of the LOGIS instructional application?

To evaluate the ADDIE process, each phase will be subject to the following five questions based on (Gagne et al., 2005, p. 355):

1. Was the rationale for including or omitting this phase clear?
2. Are all required elements of this phase accounted for?
3. Was this phase temporally cost effective to complete?
4. Are the products from this phase observable and measurable?
5. Are the products from this phase consistent with the overall goal of the instructional application?

These questions are concisely answered at the end of each phase and the results are presented in this chapter.

2. Is LOGIS an effective form of instruction?

This research question is addressed in this chapter and also reported in chapter five "Conclusions". Data from the one-to-one evaluation, the small-group evaluation, and the field trial were compiled to address this research question.

Analysis Phase Report and Reflection

The Analysis Phase was very demanding in terms of time and planning. The established need for the instruction and the existence of goals provided a good starting position and reduced the workload for this phase, but the classification of the goals proved to be very difficult. The refining of goals forced the researcher to visualize each step towards the final product in fine details. While visualization is commonly used, generating a solution path of such detail was not anticipated. The detailed visualization process was necessary because it was very important to quickly establish what was possible given the available resources, time, and expertise. It was very important to project, in detail, what the requirements would be and then make adjustments based on the anticipation of problems – all before classifying the goals. The researcher decided to exclude bar graphs and semi-logarithmic graphs because in addition to the lack of course emphasis on these topics, it was hard to visualize the programming code that could implement these graphs and also accommodate cumulative and multiple base-line graphs. Based on the level of difficulty experienced during the programming stage, the decision to exclude bar graphs and semi-logarithmic graphs was in retrospect justified.

The Analysis phase not only forced the detailed visualization of the actual application, it forced a re-envisioning of the cycle of instruction, specifically the role of assessment within that cycle. The commitment to the ADDIE phases led to the use of the KSA classification scheme, and using this scheme meant that this instruction had to include assessments for knowledge, skills, and attitudes. In this context, the assessment focuses on the participant, the content, and the application. This is a departure from the norm in that most assessments are seen as important but separate parts of instruction, and in most case, the instruction refers to the content alone. This phase forced the researcher to consider assessment as an integral and indistinguishable part of the instruction cycle, and also an integral and indistinguishable part of the application development cycle. The key point is that the assessments in this study assume expanded roles because they highlight, and to an extent mediate, the relationships between the learner and the content, the learner and the application, and the application and the content.

The process of re-envisioning of the instruction cycle led the researcher to question what type of application would be suitable for this study, and again focusing on what was possible given the available resources. In this study, instruction included content, practice, and assessment, thus the conclusion was that the instructional application had to be flexible and easily updatable. Flexibility and updateability were important qualities because both the content and the application itself would be frequently modified and revised based on the results of the assessments. In this case, flexibility and updateability included the ability to add or remove rules, examples, and practice items or entire modules. This requirement led the researcher to conclude that in addition to being an instructional application, LOGIS should be viewed as an *Instructional Engine*. The term Instructional Engine is appropriate because it accurately conveys the fact that the program should deliver a "class" of instruction. The engine in this case would deliver the graphing instruction, but it would also be able to deliver other similar instruction as deemed necessary. This becomes very important when considering the decomposition of instructional concepts. If the instructional engine is to deliver instruction on a certain concept, then it must also be able to deliver instruction on the component concepts in the event that the participant needs additional help.

The decision to build an instructional engine had a significant effect on the initial conception of the program. Attempting to build an engine meant that the software program had to be as *abstracted* as possible, but still focused on the overall instructional objective. Abstracting the engine meant that the design and implementation of the engine would be as loosely coupled to the instructional content as possible. In essence, the engine would be designed almost independent of the content, allowing for increased modularity, usability and updatability. The idea of *abstraction* is not prevalent in the instructional design area but it is a staple of software programming, thus, the theoretical and practical underpinning are established.

Analysis Phase critique

1. Was the rationale for including or omitting this phase clear?

The decision to include this phase was reasonable. Planning is accepted as an appropriate first step in many tasks, and it proved to be extremely beneficial in the case of this study.

2. Are all required elements of this phase accounted for?

The required elements of the Analysis phase were all necessary and consequently they were all completed. The only caveat is that there was no need to create goals because these existed along with the original graphing instructional unit.

3. Was this phase temporally cost effective to complete?

This phase was not temporally cost effective and it consumed an unexpected amount of time. The Analysis phase was completed in slightly less than six weeks and that accounted for 18% of the total time spent on the development of the application. Although 18% is a relatively small proportion of the total time, the researcher did not anticipate that the Analysis phase would take six weeks. Most of the six weeks were spent on the instructional analysis (Analysis Component 2). The process of conducting the instructional analysis required that the researcher pay close attention to the long-term design of the application because the results of the instructional analysis would inform the future design and development of the application. In essence, extra care was taken during the Analysis phase to anticipate and as much as possible avoid future problems.

Several factors could have contributed to the excessive time taken during this phase. The significant factors appear to be based on particular attributes of the researcher and the nature of the study as opposed to being inherent to the analysis process. The two most identifiable factors are the experience level of the researcher and role of the researcher in the study, both of which, along with their implications, are discussed in Chapter 5.

4. Are the products from this phase observable and measurable?

The Analysis phase resulted in goals that were observable and measurable. In addition to concrete goals and classifications, this phase also defined the time constraints and the participant prerequisites.

5. Are the products from this phase consistent with the overall goal of the instructional application?

The products of the stage are consistent with the overall goal of the application.

Design Phase Report and Reflection

The goals generated in the Analysis phase were used as the starting point for the Design phase. They were the basis for the tutorial topics and the basis for the assessments. The activities for each lesson and the assessments specifications were also developed in this phase.

Generating the items for each tutorial proved to be very challenging. The most difficult part was the modification of the existing tutorial content to meet the required length and format. The original tutorial contained 11 tutorial sets and these were reduced to 5 tutorial sets namely: The Control And Measurement Of Behavior, The Importance Of Graphing, Behavioral Graphing Concepts, The Cumulative Graph, and The Multiple Baseline Graph. The Primer and Basic Graphing sets were not derived from the 11 original tutorials, they were created by the researcher. Table 8 is a comparison of the 11-set original graphing tutorial and the 5 corresponding LOGIS tutorial sets.

Table 8

A Comparison of the Original Graphing Tutorials and the Derived LOGIS tutorials

Attribute	Original Graphing Tutorial Set	Derived LOGIS Tutorial Set
Total number of sets	11	5
Total number of frames	357	230
Total number of words	11077	7689
Average number of frames per set	32.45	46
Average number of words per set	1007	1537.80
Average number of words per frame	31.03	33.43

Table 8 shows the increases in frame density and word density, confirming the fact that although the number of tutorials sets decreased from 11 to 5, the density of the content increased as exemplified by the increase of average number of frames per set from 32.45 to 46 frames.

Table 9 shows the final distribution and correct sequencing of the LOGIS tutorial sets. The Basic Graphing tutorial was positioned after the Primer because the researcher felt that basic graphing should be completed first. This was logically sound, and made the programming of the application slightly easier.

Table 9

The Number of Frames in the LOGIS Tutorials and Practice Tasks

Tutorial Set Name	Number Of Frames	Number Of Items In The Practice Task
Primer	26	5
Basic Graphing	27	13
The Control And Measurement Of Behavior	26	-
The Importance Of Graphing	46	-
Behavioral Graphing Concepts	84	-
The Cumulative Graph	42	15
The Multiple Baseline Graph	32	44

During the Design phase, the total number of tutorial sets was deemed more important than the length of individual tutorial sets. In retrospect, the rationale behind placing high importance on the number of tutorial sets was flawed. During previous semesters, students who completed the original 11-set graphing tutorial reported frustration at the number of tutorial sets (11 sets) that they were required to complete. The researcher considered the data and concluded that it was more important to minimize the number of sets, at the expense of the length of each set. This did not appear to work because participants ultimately reported frustration at the number of frames in each tutorial set, especially the Behavioral Graphing Concepts set which had 84 frames. These results are reported in this chapter

The 11 original graphing tutorial sets had a combined total of 357 frames, and the LOGIS tutorial sets and practice tasks had a combined total of 360 frames. Although the number of tutorial sets in LOGIS is less than the number of tutorial sets in the original tutorial, the LOGIS tutorial sets contained more total frames (including practice frames) than the original tutorial. This significance of this situation was not evident until the development phase and by that time it was too late to make large scale changes because of the beta testing deadlines. Table 10 shows a comparison between the original graphing tutorial and the LOGIS tutorial.

Table 10

A Comparison of the Original Graphing Tutorials and the LOGIS Tutorials

Attribute	Original Graphing Tutorial	LOGIS Tutorial including Practice Tasks	LOGIS Tutorial excluding Practice Tasks
Total number of sets	11	7	7
Total number of frames	357	360	283
Total number of words	11077	11699	9251
Average number of frames per set	32.45	51.43	40.43
Average number of words per set	1007	1671.29	1321.57

Attribute	Original Graphing Tutorial	LOGIS Tutorial including Practice Tasks	LOGIS Tutorial excluding Practice Tasks
Average number of words per frame	31.03	32.50	32.69

The original graphing tutorial had more sets than the LOGIS tutorial, but the latter had more frames per set and more words per frame. The relatively high frame density is consistent when the practice tasks are included (51.43 frames per set) and when the practice tasks are excluded (40.43 frames per set). This situation directly resulted in participant frustration, which is reported in this chapter.

The Primer tutorial and the corresponding practice task were created during the Development phase, not the Design phase. This was not an oversight because the necessity of a Primer was evident from the beginning of the study. The Primer was not initially perceived as instruction, thus it was developed separately from the rest of the tutorial sets. In retrospect, this was an error because all tutorial sets should have been viewed as instruction and consequently the Primer should have been created with a focus on systematic development, similar to the other tutorials.

Design Phase Critique

1. Was the rationale for including or omitting this phase clear?

The Design phase was critical because it resulted in the formation of most of the instructional content. This phase could not have been omitted.

2. Are all required elements of this phase accounted for?

The required elements of the Design phase were all completed.

3. Was this phase temporally cost effective to complete?

This phase took an significant amount of time, but this was expected. The Design Phase was completed in about nine weeks, which was approximately 27% of the total time. Most of the time in this phase was spent deconstructing the original tutorial sets and assembling the sets that would be used for LOGIS. The researcher expected that a few problems to occur when converting instruction from one form to another, but no problems occurred. The creation of new content and the modification of existing content were expected to be labor intensive and this proved to be the case. This phase was temporally cost effective and it would be difficult to argue otherwise given the fact that the expectation was that phase would take a long time.

4. Are the products from this phase observable and measurable?

The Design phase resulted in the detailed documentation of the goals, the creation of the tutorials sets, the finalization of the instruction sequence, and the creation of assessment specifications. The products are all observable and measurable.

5. Are the products from this phase consistent with the overall goal of the instructional application?

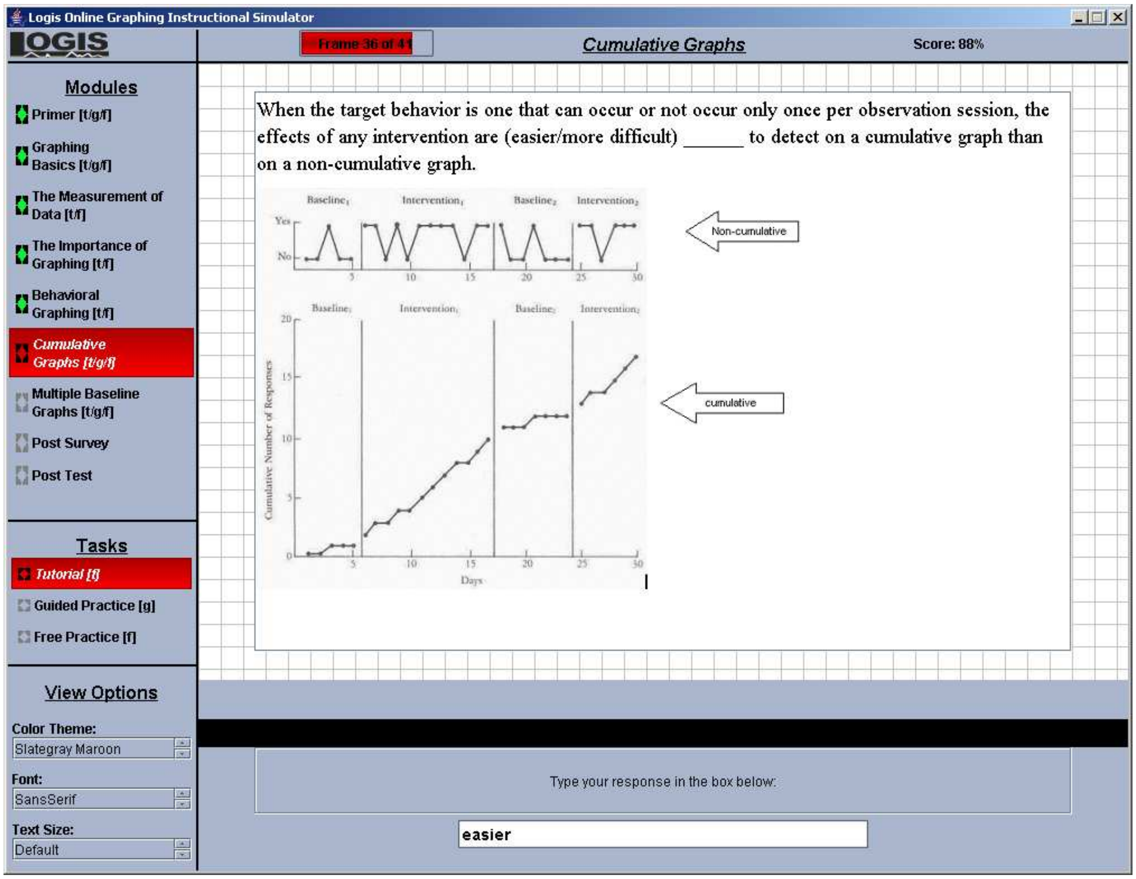
The products of the stage are consistent with the overall goal of the application.

Development Phase Report and Reflection

The interface and the application

The final version of the LOGIS interface does not differ significantly from the proposed design, primarily due to the extensive planning during the Analysis phase.

Although the interface design changed from concept to implementation, the changes were superficial and were largely the result of incorporating new features into the application. The basic interface components and their interactions were maintained from the proposed design (see Figure 4) to the completed application. Figure 11 shows the final version of the LOGIS interface.



When the target behavior is one that can occur or not occur only once per observation session, the effects of any intervention are (easier/more difficult) ______ to detect on a cumulative graph than on a non-cumulative graph.

Color Theme: Slategray Maroon

Font: SansSerif

Text Size: Default

Type your response in the box below:

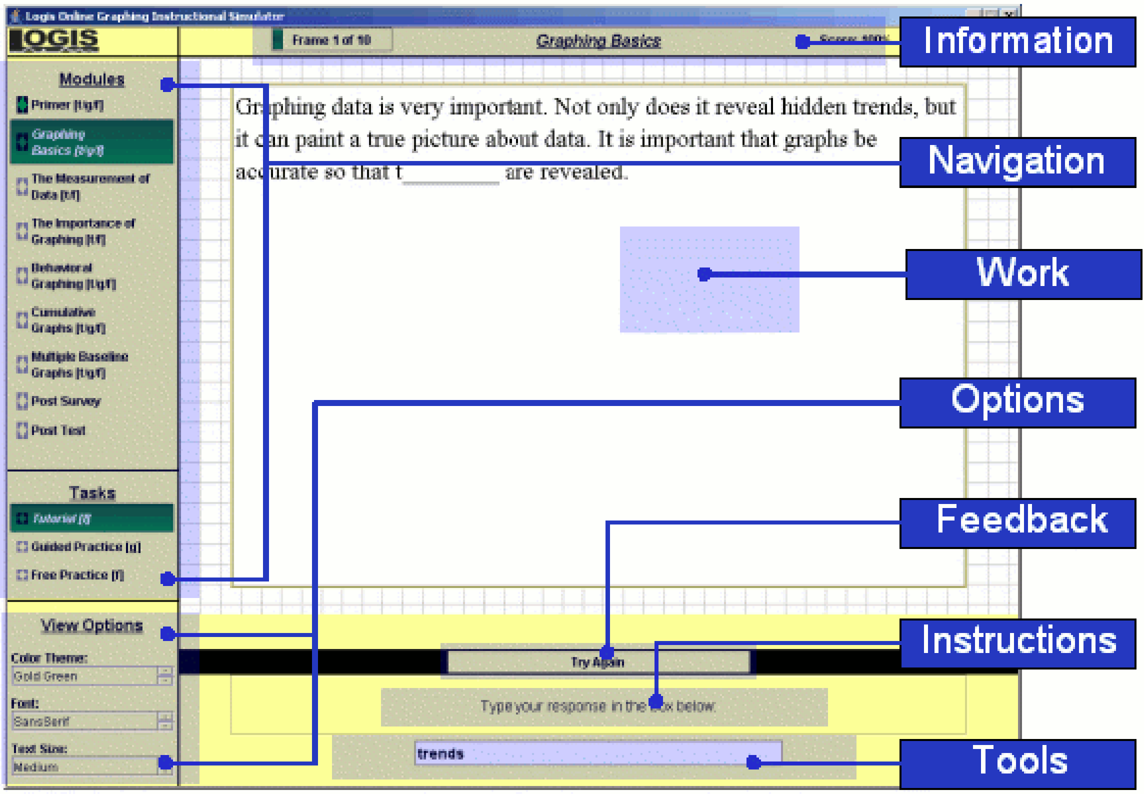
easier

Figure 11. A view of the final version of the LOGIS interface.

During the Development phase it became evident that the terms that were being used were insufficient. The references to tutorials and practice tasks became cumbersome during the programming of the application. At that point, the researcher decided to change some of the terminologies to better reflect the components of especially the interface. The terms *Modules* and *Tasks* were introduced as “containers”, where Module was used as a high-level container for all possible Tasks. This meant that the LOGIS application could now be described as having modules containing tasks, and the tasks were a combination of tutorials and practices. There were 12 modules: Pretest, Pre-survey, Graphing Proficiency Test, Primer, Graphing Basics, The Measurement of Data, The Importance of Graphing, Behavioral Graphing, Cumulative Graphs, Multiple Baseline Graphs, Post-survey, and Posttest. The tutorials created during the Design phase were embedded within their respective modules and were paired with at least one form of practice. The advantage of using the module concept as a high-level container was that it reflected the theme that everything within LOGIS is a part of the instruction. Using the simple concept of modules allowed non-traditional forms of instruction, for example tests and surveys, to be implemented on the same conceptual and physical level as traditional forms of instruction, for example tutorials.

Practice is one of the fundamental principles of LOGIS. During the Development phase, the researcher decided that the participants should have the ability to practice graphing even if the module they were completing did not have a Guided Practice task. This led to the development of the Freelance Practice. The Freelance Practice, labeled Free Practice on the interface, gave the participant the opportunity for unguided practice during any non-survey or non-test module. The Freelance Practice task was available after the participant completed a tutorial, but if a Guided Practice task was present, then the Freelance Practice became available after the Guided practice task was completed. The purpose of the Freelance Practice task was to provide unguided practice after the completion of each module's formal instruction.

The final LOGIS interface was comprised of seven areas: Information, Navigation, Work, Options, Feedback, Instructions, and Tools (see Figure 12).



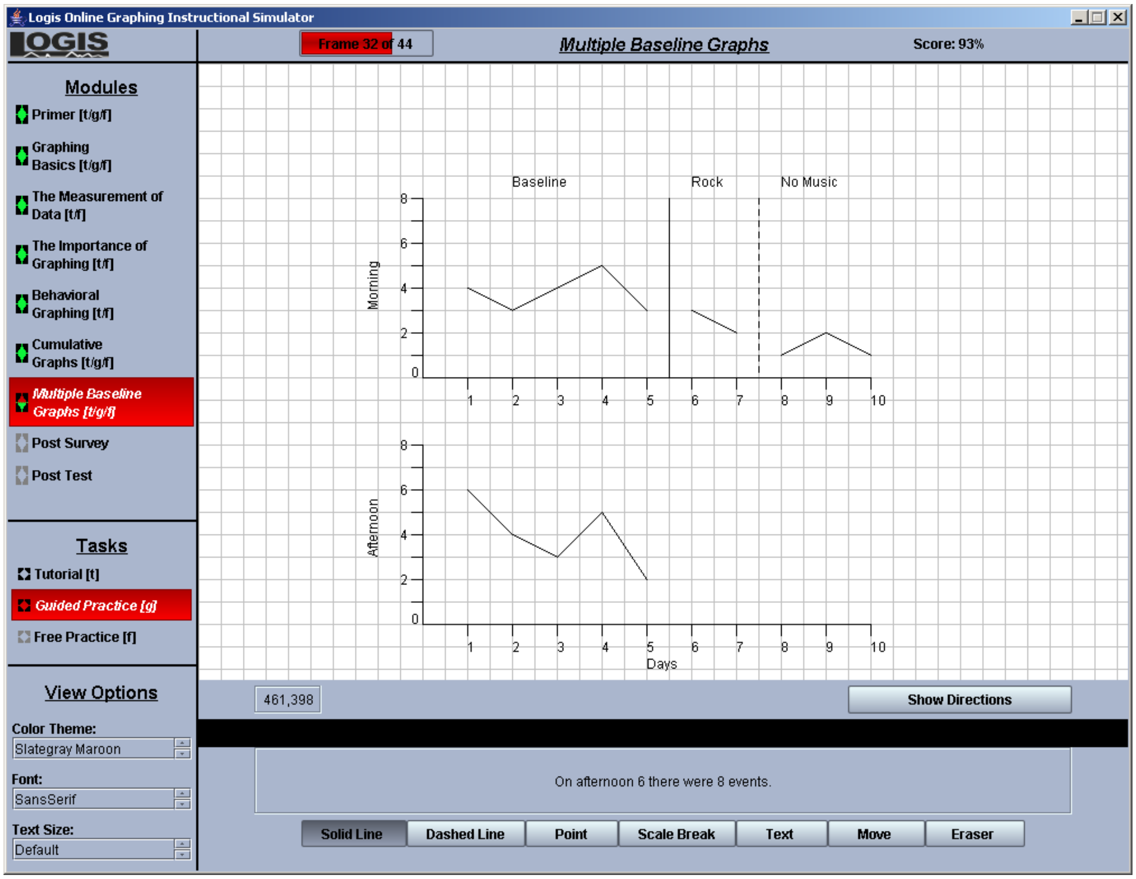
Graphing data is very important. Not only does it reveal hidden trends, but it can paint a true picture about data. It is important that graphs be accurate so that t________ are revealed.

Type your response in the box below.

trends

Figure 12. The significant areas of the LOGIS interface.

The Information area displayed the current module, the score on the current task, and the frame count for current task. The frame count, which is an indication of progress, was displayed both textually and graphically. The Navigation area displayed the modules and tasks. The Work area displayed the contents for the tutorials, the tests, the surveys, and the initial instructions for each task. The Work area also doubled as the graphing workspace during the Guided Practice and Free Practice tasks. The Options area allowed the participant to change the color theme, the font style, and the text size. The Feedback area provided textual feedback to the participant if a response was incorrect, prompting the participant to “Try Again”. The Instructions area provided explicit directions on what should be done at that point in the instruction. The Tools area provided the participant with the tools to respond. Tutorials and tests required a text box for alphanumeric input (see Figure 12) and the practice tasks required graphing tools for graphical input, as visible in Figure 13. Surveys required a slide bar that corresponds to numeric input (see Figure 14).



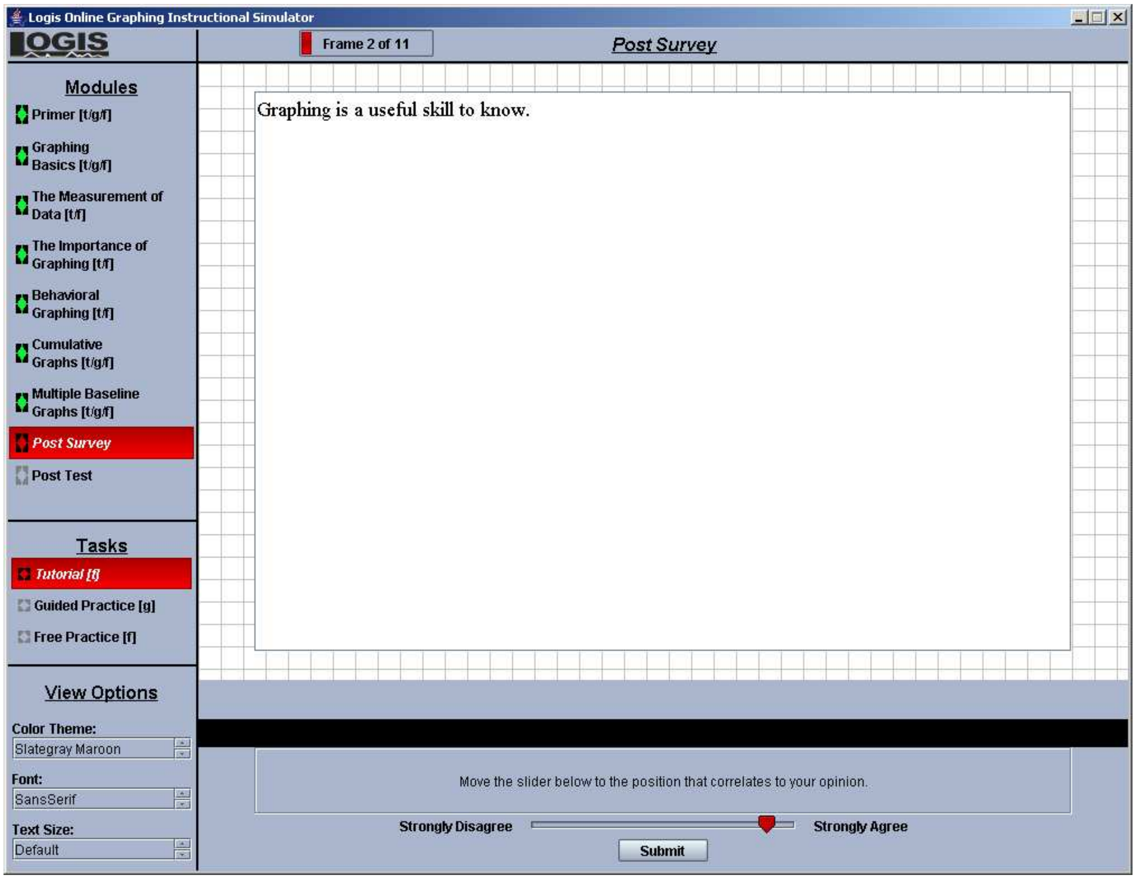
Color Theme: Slategray Maroon

Font: SansSerif

Text Size: Default

On afternoon 6 there were 8 events.

Figure 13. A view of the LOGIS interface showing the grid and graphing tools.



Graphing is a useful skill to know.

Move the slider below to the position that correlates to your opinion.

Color Theme: Slategray Maroon

Font: SansSerif

Text Size: Default

Figure 14. A view of the LOGIS interface showing a survey item and the slide bar.

The functionality of LOGIS can be best described by looking at the major features that were implemented:

The application environment

- LOGIS was developed as a stand-alone Java application. The original specification called for a Java Applet, but during the Development phase it became apparent that browser restrictions and the available browser space would cause significant problems. Programming LOGIS as a stand-alone application allowed the researcher to control every aspect of the display environment.
- As defined in the proposed specifications, a Microsoft SQLServer database was used as the container for both input and output data.

- The LOGIS interface was designed for a minimum screen resolution of 1024 pixels wide and 768 pixels high.
- A rudimentary form of versioning was used to ensure that participants were always using the most current form of LOGIS. After a successful login, the current application version was queried from the database and then compared to the participant's version. If the versions did not match, the application halted and then a message and URL address were displayed. The message prompted the participant to download the current LOGIS version at the URL address.

Login procedures

- Login was based on a username and password. In this study, the username was the participant's university login identification and the password was the participant's student number.
- The participant's completed tasks were queried from the database after a successful login, after which the appropriate links and icons in the navigation area were enabled. This session management ensured that participants could return to their current module in the event that they needed to restart the application.
- Participants had access to only the current module and previously completed modules. This ensured that the modules were completed sequentially.

Options

- The color theme option controlled the background color, the text color, and the highlight color for buttons and text. Changing the color theme applied the selected color theme to the entire interface.
- The font option allowed the participant to select among Monospaced, Serif, and

SanSerif fonts. These fonts had differing sizes and letter spacing but they did not decrease readability. This option gave the participant an opportunity to use their preferred font.

- The text size option allowed the participant to change the size of the text in the Work Area and in the Instructions Area. These areas are labeled in Figure 12.

Tutorials

- When a module was selected by a participant, the tutorial task for that module automatically started.
- Exhibits were originally envisioned to be in a tabbed (embedded) window separate from the tutorial frame, but this was changed in an effort to make the LOGIS interface less demanding on the participant. The exhibits were displayed in the Work Area beneath the frame text, as seen in Figure 11.
- Participants used a text box to respond to frames. The text input was controlled using regular expressions. This prevents participants from entering illegal and possibly harmful text.
- Each frame required a response. Feedback in the form of “Try Again” was presented in the Feedback Area (see Figure 12) if a response was incorrect. If the participant had exhausted the maximum number of response attempts allowed for that frame, the correct answer was displayed and then the learner was required to type the correct answer to proceed to the next frame.

Practice tasks

- The participant read the current instruction from the Instructions Area, and then used the mouse to draw or place shapes on the grid (see Figure 13).

- Participants could draw solid or dashed lines; place points, text, or scale breaks; move any drawn object; erase specific objects; or they could erase all the objects on the grid.
- All objects drawn by the participant were initially colored black. If the object was incorrect, feedback in the form of "Try Again" was presented in the Feedback Area (Figure 12) and the object's color was changed to red. If the participant had exhausted the maximum number of response attempts allowed for that frame, the correct object was displayed in colored blue as a hint. The participant then had to draw the correct object to proceed to the next frame.
- LOGIS used the location of pixels on the drawing area to determine the position of the mouse pointer. The participant, however, did not need to be precise when drawing objects. During this study the fidelity of the answers was set to 10 pixels, thus an object only needed to be within 10 pixels of the correct answer to be considered correct. For example, if the tutorial asked the participant to place a point at pixel position (100,250), correct answers would be defined by the square: (90,240), (110,240), (110,260), and (90,260). This reduced the need for precise mouse-manipulation on the part of the learner.
- To avoid repetition, some graph elements were automatically populated. For example, the participant was asked to place the first two hatch marks on an axis, after which the rest of the hatch were automatically populated.
- Freelance Practice did not have any tutorial or practice task associated with it. This area provided the participant with the opportunity to interact freely with the application. The Freelance Practice did not evaluate the participant's responses,

consistent with the idea of free interaction.

Tests and surveys

- A perfect score on the Graphing Proficiency test exempted the participant from the Graphing Basics module, although the Graphing Basics module could still be completed if the participant wished.
- The current score was not displayed when the participant was taking a test.
- The survey slide (see Figure 14) presented an opportunity to collect the participant’s attitude with a high level of precision. The extreme numerical values of the slide were 1 and 1000 and they were labeled strongly disagree and strongly agree respectively. The slide was centered by default and this corresponded to a numerical value of 500.
- Participants had to confirm survey responses that were truely neutral. If the participant submitted a neutral response (the slide is in the middle position), a confirmation dialog box would popup and require that the participant confirm the neutral selection. This prevented participants from repeatedly clicking the submit button without actually giving their opinions.

The original development plan for LOGIS (see Figure 9) was significantly changed during the Development phase. Figure 15 shows the actual development plan that was used.

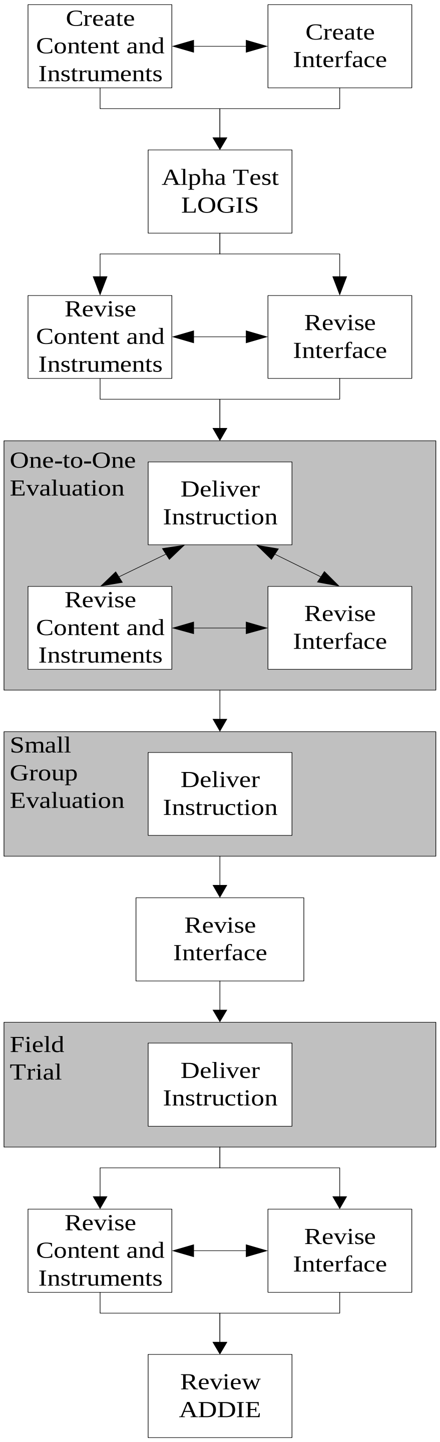


Figure 15. A flowchart of the actual development process.

The rationale behind changing the development plan was focused on the definition of evaluation. The originally planned program evaluation of LOGIS included one-to-one evaluation, small-group evaluation, and a field trial. These evaluations were slated to occur during the Evaluation phase, but after careful consideration no clear reason emerged as to why these three evaluations could not be implemented in the Development phase and repurposed as a form of formative evaluation. The benefits of this change in perspective seemed to heavily outweigh any possible negative effects. Firstly, the original beta tests were fully contained within the three evaluations. The new plan would provide three opportunities for evaluation instead of the proposed two beta tests opportunities. Secondly, the new plan more accurately reflected the idea that although LOGIS is a singular application, its interface and its content could be seen as independent individual components. In essence, the new plan reflected the idea of LOGIS as an instructional engine, where the application and the interface are somewhat decoupled from the content. Thirdly, the new plan better modeled the correct sequence of development events. Unlike the original plan, the new plan clarified that the creation and revision of the content and the interface occurred simultaneously where the content affected and was affected by the interface. Finally, the new plan highlighted the high level of interaction between the researcher and the participants that should occur early in beta testing. In this case, the one-to-one evaluation presented an opportunity to address issues immediately, allowing instant revision of the content and the interface while the application was being tested.

The population described in “Evaluate Component 2” in Chapter 3 formed the pool of students who were eligible to complete the evaluation tasks, and 47 students volunteered to evaluate LOGIS. The participants were divided into 3 groups. The one-to-one evaluation group contained 3 participants, the small-group evaluation group contained 13 participants, and the field trial group contained 31 participants. Participants were assigned to groups based on their current course grade and the specifications outlined in the Evaluation phase of the proposed development (chapter 3). In certain cases, however, participants were placed in a particular group because it was more convenient for the participant. This did not present a problem because the group did not need to be rigorously defined because the purpose of this evaluation was to refine the instruction, not to experimentally investigate the effect of the instruction.

Tutorial Development

The tutorial tasks were developed as proposed in the Design phase, with a small but significant change to the manner in which frames were presented. During the programming of the application and after consulting with the content expert, a full professor who specializes in the content area and is responsible for the behavioral courses, the researcher decided to introduce a feature that would guarantee that each participant responded correctly to each frame. Under normal conditions, each frame is assigned a number of attempts that determines how many mistakes the participant can make before the frame is considered wrong. For example, a frame that is assigned two attempts allows one mistake, and then the learner must make the correct response on the second attempt or no points are earned for that frame. If the second attempt is wrong, the correct answer is displayed and then the next frame is presented. The new feature, however, required that the participant enter the correct response even if the second attempt was wrong thus the instruction only proceeded when the participant entered the correct response. Under the new scheme, when the maximum number of attempts was exhausted the correct answer was displayed but the participant had to type the correct response to continue. The benefit of this feature was that the learner came in contact with the correct response on each frame.

Under normal circumstances, each frame of each task would be examined for revision based on the percentage of participants who responded correctly to that frame. Frames that did not have a high percentage of correct responses, for example frames that were answered correctly by less than 90% of the participants, would be examined for errors and revised. During this study, however, the participant had to respond correctly to each frame before instruction could continue thus in essence every participant answered every frame correctly, although points were not earned if the participant exceeded the maximum number of attempts. This design decision was pedagogically sound but it made the analysis of the resulting data difficult. It was now difficult to determine if many participants entered wrong responses or one participant entered many wrong responses for a given frame. Given this situation, each frame had to be examined individually to determine if the frame contained errors that affected many participants or if the number of wrong responses were from a few participants.

The tutorial task data, which excluded the practice data, were examined in a systematic manner starting with frames that contained the highest number of wrong responses, that is, frames with high error rates. Table 2 in Appendix L (Table 1 is the legend for Table 2) shows the number of wrong responses for each frame of each task and is the basis for the analysis of the tutorial task data. Frames that had high error rates were examined for patterns that might indicate why participants entered wrong answers. If no discernable pattern could be seen, the frame was marked for revision or deletion.

The reasons why certain frames had high error rates can be divided into two categories, participant generated and frame generated. Participant generated errors were errors caused by the participant. Examples of these errors included grammar mistakes, spelling mistakes, and using the wrong tense. Frame 3 of the Basic Graphing tutorial task (task 6 in Appendix L) is an example of a grammar mistake where some participants entered the response “two axis” as opposed to the correct response “two axes”. Spelling mistakes were the single most common reason why participants got frames wrong. Words like perpendicular, labeled, and cumulative, are just a few of the words that participants regularly misspelled. The second category, frame generated, refers to errors resulting from the frame’s content. In most cases, participants responded incorrectly to the frames that were poorly primed or excessively complex. The same result occurred when frames contained grammar, spelling, or content errors. Frame 6 of the Behavioral Graphing Concepts tutorial task (Task 10 in Appendix L) is an example of a frame that had no discernable pattern of errors because it was poorly primed, thus it needs to be revised or removed from the tutorial task.

Knowledge Assessment Development

The knowledge assessment was created using the methodology outlined in chapter three. The final version of the assessment contained 40 items, 20 alternate-choice items, and 20 short-answer items (see Appendix M). As planned, 70% of the items were higher order items. Table 11 shows the distribution of the items.

Table 11

The Distribution of the Knowledge Assessment Items

Tutorial	Weight	Short Answer	Alternate Choice
The Control And Measurement Of Behavior	20%	4	4
The Importance Of Graphing	20%	4	4
Basic Graphing	0%	0	0
Behavioral Graphing Concepts	20%	4	4
The Cumulative Graph	20%	4	4
The Multiple Baseline Graph	20%	4	4

The knowledge assessment items were created by the researcher then examined by a member of a research consultancy group at the researcher's university. The consultant made suggestions regarding the wording of certain items, and recommended that multiple-choice items be excluded because of time constraints. Based on the consultant's recommendation, no multiple-choice items were included in the knowledge assessment.

The items were tested during the one-to-one evaluation sessions and mistakes including typographical errors and grammatical errors were corrected before the statistical analysis of the data. This meant that participants were not penalized for spelling mistakes or simple grammatical errors. The small-group evaluation and field trial evaluation were conducted after the one-to-one evaluation, and they included the refined 40 items. The items, however, were not changed during the small-group or field trial evaluations.

Of the 47 participants who participated in the study, 35 completed the final version of the Knowledge Assessment. The remaining 12 participants received assessments that were either revised or the application returned an error during their assessment. These 12 participants included the three participants chosen for the one-to-one evaluation, and the next 9 participants who started the application from the small-group evaluation group. Table 12 shows data that describes the overall performance on the posttest.

Table 12

A Summary of the Statistical Data from the Posttest

Source	*n*	Min	Max	*Mdn*	*M*	*SD*	skewness	kurtosis
Posttest	35	37.50	85.00	62.50	62.00	11.08	-0.44	0.27
							SE=0.40	SE=0.78

The posttest scores were approximately normally distributed based on the values for skewness and kurtosis. As a guide, if the skewness and kurtosis values fall within twice their Standard Error ranges from positive to negative, then the distribution is considered symmetrical with respect to normal distribution. Twice the skewness standard error (0.40) resulted in a range of -0.80 to 0.80. The posttest skewness (-0.44) fell within the -0.80 to 0.80 range, thus the distribution was symmetrical in terms of the skewness. The kurtosis (0.27) was within twice its standard error's (0.78) range of the -1.56 to 1.56, thus the distribution was neither overly peaked nor flat. Figure 16 is a histogram of the posttest scores showing the distribution of the grades.

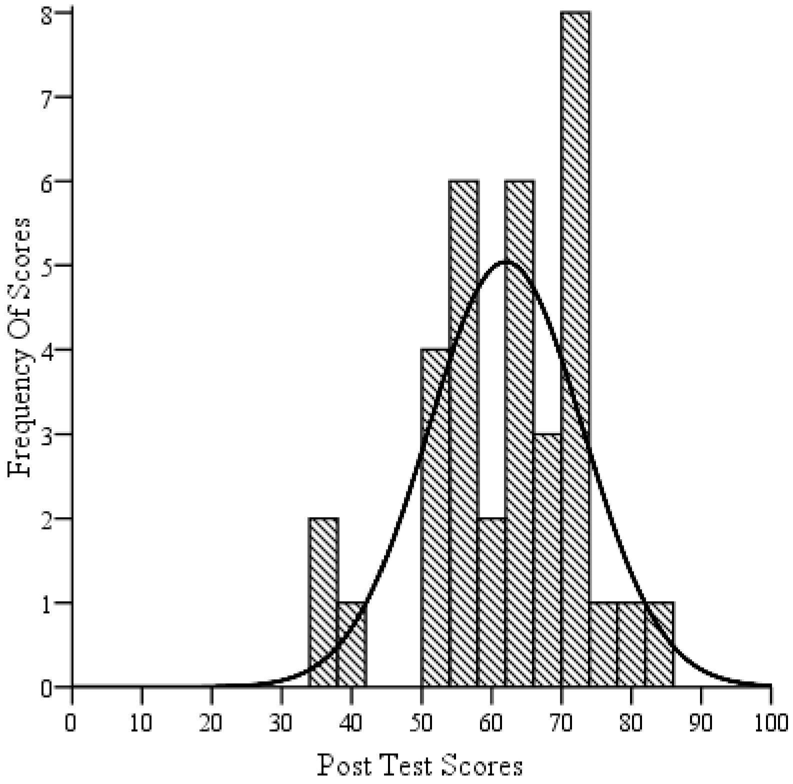


Figure 16. A Histogram of the posttest scores showing the distribution of the scores.

The posttest mean (*M* = 62.00, *SD* = 11.08) indicated that performance on the posttest was generally unsatisfactory. Three extremely low scores, 37.50%, 37.50%, and 40.00%, decreased the overall mean, but these outliers did not explain the overall poor scores. The low posttest scores were also a source of concern because half the posttest was Alternate-Choice items consequently it was possible that participants received higher than normal scores by guessing the correct answers on the Alternate-Choice items. The posttest was a Criterion Referenced test and normally a highly negatively skewed distribution is the preferred outcome. Considering the nature of the test and the test items, a low mean score may indicate a failure of the instruction, a failure of the testing instrument, or a combination of both. In this case, however, it must be noted that this was only the initial stage in developing a valid and reliable Knowledge Assessment. Under real-world conditions, the 40-item pool would be reduced using, for example, the Item Difficulty Index and the Item Discrimination index, thus the resulting data would be a more accurate reflection of the participants' mastery of the instruction.

The internal consistency reliability, the Cronbach's Alpha, of the posttest was .73. Appendix N describes the data that was generated from the posttest. In addition to the Item Difficulty Index, the Point-Biserial Correlation (*r*pb) was used to analyze each test item. The Point-Biserial Correlation (Pearson Product Moment Correlation) indicates the relationship between a participant's score on an individual item (dichotomous) and the participant's overall test score (continuous). It is normally used to explore how the item discriminates among participants, and this is normally a useful statistic for tests designed to spread participants, for example, Norm-Referenced tests. In this case, however, it was used with the Item Difficulty Index to provide additional insight into items that may problematic. Items with negative Point-Biserial Correlations, for example -.08 for item number 5 (see Appendix N) may not be measuring the correct construct and should be examined and revised or eliminated. Based on the Item Difficulty Indexes, the Point-Biserial Correlations, and the nature and wording of the items themselves, items 5, 6, 13, 18, 24, 36 in Appendix N were the first items selected for review. In addition, the selected items had Corrected Item-Total Correlations of -.16, -.13, .30, .28, .33, and .11 respectively, making them prime candidates for revision because they appeared to be negatively or weakly correlated to the overall test. It must be noted that in addition to having the largest negative Point-Biserial Correlation and Corrected Item-Total Correlation, item number 5 also had the greatest effect on the statistical reliability of the posttest. If item number 5 was removed, the Cronbach's Alpha would increase from .73 to .74. Although some items have moderate difficulty indexes, for example items 1, 3, 8, 15, and 31, they could be useful because they discriminate between individual participants.

The Knowledge Assessment was administered as the pretest and as the posttest, where the posttest immediately followed the instructional content. Of the 35 participants who completed the final version of the Knowledge Assessment, 29 participants completed matching pretests and posttests, that is, the pretest and posttest had identical items. The remaining 6 participants completed posttests that were edited after the pretest was completed. Table 13 shows data that describes the overall performance on the pretest and posttest (matching items) by the 29 participants.

Table 13

A Summary of the Data from Participants Who Completed Matching Pretests and Posttests

Source	*n*	Min	Max	*Mdn*	*M*	*SD*	skewness	kurtosis
Pretest	29	27.50	57.50	42.50	43.36	8.22	-0.22	-0.88
							SE=0.43	SE=0 .85
Posttest	29	37.50	85.00	65.00	62.41	11.96	-.54	.01
							SE=0.43	SE=0.85

The Cronbach's Alpha of the pretest was .50 and it was .73 for the posttest. The pretest had a low reliability alpha, but that value has to be considered with caution because it could have been the result of random guessing. Using twice the skewness and kurtosis standard error ranges as guides, the pretest and posttest scores were determined to be normally distributed. The skewness of the pretest (-0.22) was between -0.86 and 0.86, and its kurtosis (-0.88) was between -1.69 and 1.69. The skewness of the posttest (-0.54) was between -0.86 and 0.86 and its kurtosis (0.01) was between -1.69 and 1.69. Based on the standard error ranges, both distributions do not significantly depart from symmetry with respect to skewness and kurtosis. Figure 17 shows the histograms of the pretest and the posttest scores of the 29 participants and provides an opportunity to compare the participants' performance on the tests.

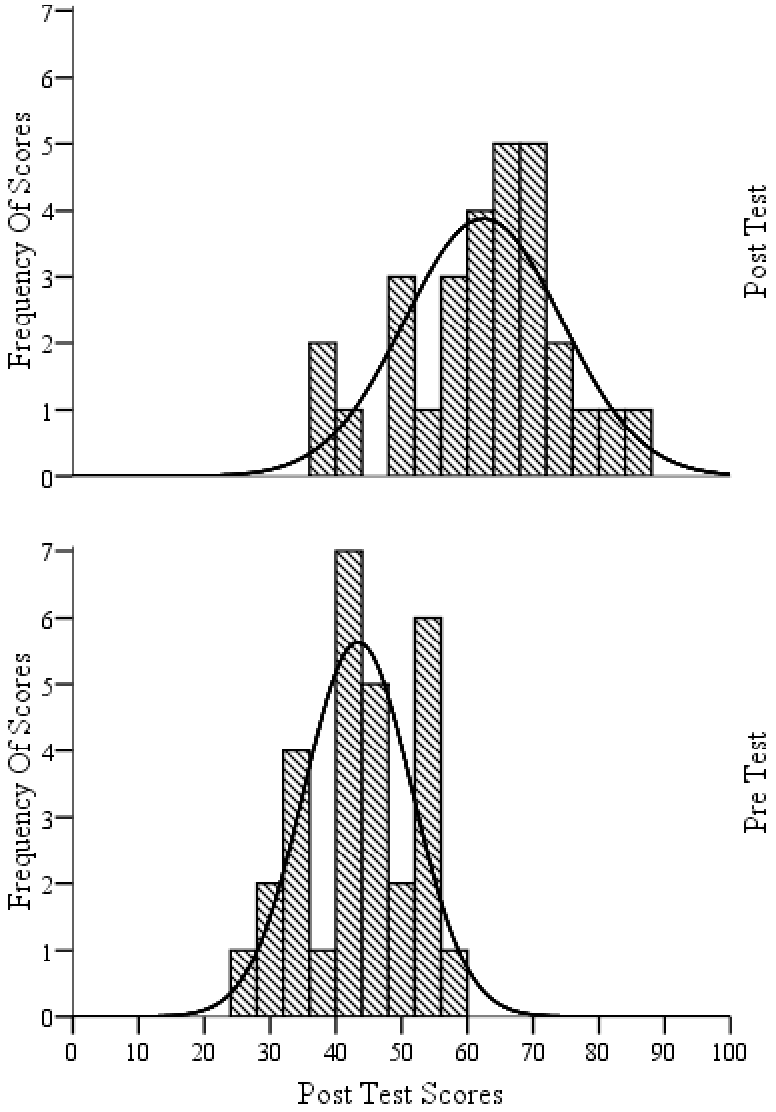


Figure 17. Histograms of the pretest and posttest showing the distribution of the scores.

Participants were expected to perform poorly on the pretest but although the mean (*M* = 43.36, *SD* = 8.22) was low, it was higher than expected. The possibility of inflated scores should have been anticipated because similar to the posttest, the pretest was 50% Alternate-Choice items, therefore, it is possible that participants received higher than normal scores on the pretest simply by guessing the correct answers on Alternate-Choice items. Similar to the posttest, the pretest was only an initial screening of test items and it was expected to return more consistent results as the items were revised.

The 29 participants who completed the matched pretest and posttest were a subset of the 35 participants who did not experience problems while completing the posttest. The posttest data of the 29 participants were very consistent with the data from the overall 35-participant group. The outliers, visible in Figure 18, probably had great influence in the overall mean score of the 29-participants subset because of the smaller sample size. These outliers did not entirely explain the low mean (*M* = 62.41, *SD* = 11.96) but they provided further evidence that the Knowledge Assessment might need refining.

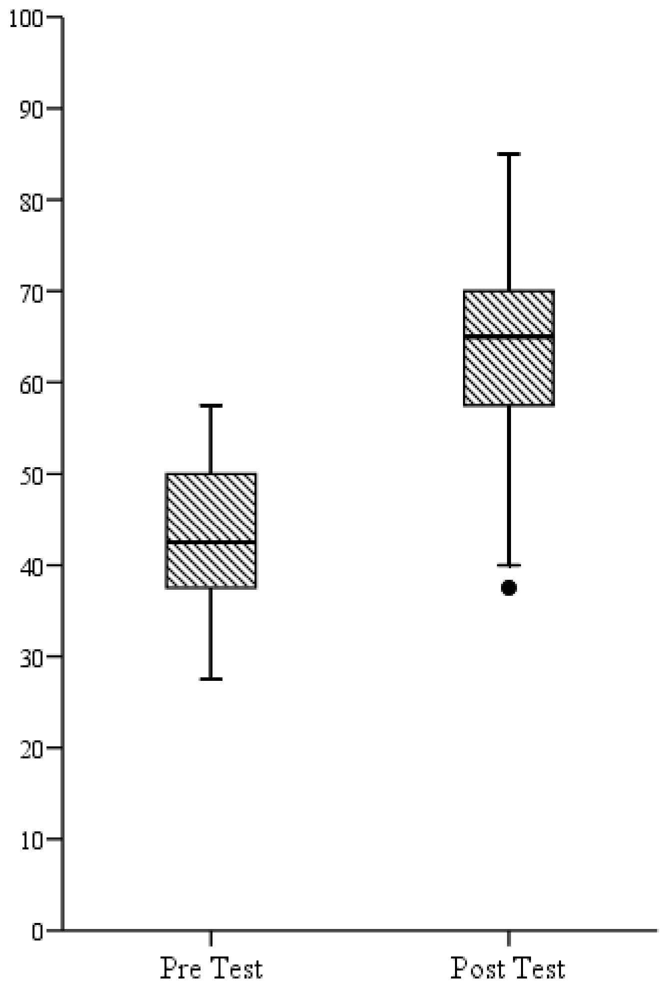


Figure 18. Boxplots of the scores for the 29 participants with matched pretests and posttests.

The means for the pretest and the posttest were compared using the paired-samples *t* test procedure. This procedure assumes that the mean differences are normally distributed and that each variable pair is observed under the same conditions, although the variances of each variable can be equal or unequal. A visual inspection of the histograms in Figure 17 and the boxplots in Figure 18 revealed that the data were normally distributed. The normality of the scores was confirmed by the skewness and kurtosis, where each was within the bound of twice its standard error. Each pretest-posttest pair did not occur under exactly the same conditions. Some participants completed the pretest and posttest locally in the computer laboratory while others completed the tests remotely. This difference is significant but acceptable because this is an initial analysis that will be used to refine the Knowledge Assessment.

The paired-samples *t* test revealed significant differences in the pretest scores and the posttest scores, $t(28) = -10.30$, $p < .01$, $\alpha = .01$ and a large effect size $d = 1.91$. The effect size (Cohen's *d*) was the mean difference between the item pairs divided by the standard deviation of the item pairs. The one-tailed p-value ($p < .005$) indicated that the mean posttest score ($M = 62.41$, $SD = 11.96$) was significantly higher than the mean pretest score ($M = 43.36$, $SD = 8.22$). The data provided an opportunity for an initial review of the usefulness of the LOGIS application. There was an increase in the scores from pretest to posttest on the Knowledge Assessment but LOGIS cannot be assumed as the cause of the increase. What the data does suggest, however, is that the instructional application was on the right track. A 10% increase from pretest to posttest was established as the criterion for educational significance, thus an increase of 19.05% confirms that the results were educationally significant within the defined criterion. Although the results were significant, they did not represent acceptable performance because the posttest mean of 62.41% did not represent a passing grade. These results should be considered within the context that the evaluation was not experimental and that the assessment instrument was in the initial stages of being refined. It is expected that as the instruction and the assessment instrument improve, the results will be more significant and useful.

Skills Assessment Development

There were significant differences between the proposed development of the skills items (Development Component 2 in chapter three), and the actual development of the items. The initial steps of the modified Crocker and Algina (1986, p. 66) 10-step process were implemented. The primary purpose and the representative behaviors were identified, the specifications were generated and the items were created then reviewed by the content expert. The Skills Assessment originally contained four items, two items required that the participant draw Cumulative graphs, and the other two items required that the participant draw Multiple Baseline graphs. After considering the time it would take for participants to complete the Skills Assessment, the decision was made to remove one type of each graph. The resulting Skills Assessment had two items, one Cumulative graph item and one Multiple Baseline graph item (see Appendix O).

The Skills Assessment was only tested during the one-to-one and the small-group evaluations, and it was not delivered before the instruction as a pretest. The decision to forgo giving the Skills Assessment as a pretest, stemmed from the concern that the post-Skills Assessment results would be significantly influenced by a Skills pretest. The original flowchart of this study would have accommodated a pre-Skills test because sufficient time would have elapsed between the pretest and the posttest to minimize the testing threat. Changes in the overall design of the study and time constraints prevented the optimal pre-post design.

Validity and reliability studies were limited to examination by the content expert, thus the items did not change during the different evaluations. The original study design called for multiple raters to be used and the Intraclass Correlation Coefficient (ICC) to be calculated. In actuality, only one rater was used, and the ICC was not calculated. The change was made based on the advice of the content expert. The rationale was that the Skills Assessment contained items that the participant was required to know after completing the instruction. Thus, the aim was not to generate skills items that would distribute the participants, rather the items reflected the standard that participants should achieve.

The Skills Assessment was graded by the researcher using the point allocation scheme proposed in Table 7 and a rubric. The rubric was developed by the researcher and it consisted of 12 points that were common to both the Cumulative and the Multiple Baseline graphs, and 2 points that were specific to the Multiple Baseline graph. The rubric covered the following points:

1. The ratio of the X and Y axes lengths is appropriate.
2. The X-axis is scaled appropriately.
3. The Y-axis is scaled appropriately.
4. The units on the X-axis are labeled correctly.
5. The units on the Y-axis are labeled correctly.
6. The X-axis is labeled correctly.
7. The Y-axis is labeled correctly.
8. The data points are plotted correctly.
9. The data points are demarcated correctly.
10. The data points are connected using straight lines.
11. The graph has an appropriate title.
12. The conditions are labeled correctly (Multiple Baseline graph only).

13. The conditions are separated correctly (Multiple Baseline graph only).

Consistent with the proposed design, participants were graded based on the rubric after which the score was converted to a percentage. A total of 16 participants completed the Skills Assessment. These participants were from the one-to-one evaluation group and the small-group evaluation group. Table 14 is a summary of the 16 participants' grades on the Skills Assessment.

Table 14

A Summary of the Statistical Data from the Skills Assessment

Source	*n*	Min	Max	*Mdn*	*M*	*SD*	skewness	kurtosis
Cumulative	16	39.39	93.94	77.27	75.76	13.78	-1.12 SE=0.56	2.02 SE=1.0
Multiple-Baseline	16	0.00	97.44	67.95	61.21	29.35	-0.59 SE=0.56	0.46 SE=1.09

The overall performance on the Cumulative Graph task was average (M=75.76%, SD=13.78), while performance on the Multiple-Baseline Graph task was below average (M=61.21, SD=29.35). The Boxplots in Figure 19 shows the distribution of the Skills Assessment grades.

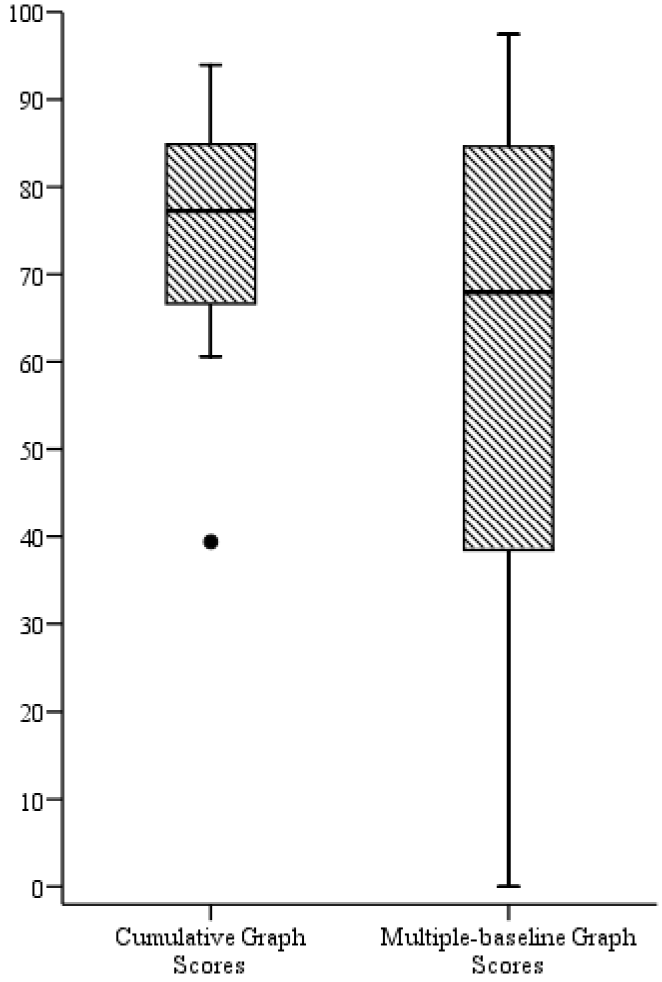


Figure 19. Boxplots of the Skills Assessment scores

In an effort to determine the specific areas where participants experienced difficulty, the average score for each item on each rubric evaluation point was calculated. Table 15 shows the average score for each item on the Cumulative Graph task and the Multiple-Baseline Graph task, where 3 is the maximum points for each item.

Table 15

The Average Score on each Rubric Evaluation Point for each Skills Assessment Graph

Source	1	2	3	4	5	6	7	8	9	10	11	12	13
Cumulative	1.7	1.9	1.6	2.9	2.6	2.9	2.8	1.6	2.6	2.7	1.7	-	-
Multiple-Baseline	1.7	1.4	1.8	2.1	2.4	2.1	2.1	1.7	2.4	2.2	1.5	1.3	1.1

The averages of the items revealed several areas where participants had difficulty. Participants performed poorly on items 1, 2, 3, 8, and 11 on both the Cumulative Graph task and the Multiple-Baseline Graph task. They also performed poorly on items 12 and 13 on the Multiple-Baseline Graph task. The data revealed that participants had difficulty with the axes lengths and scales (items 1, 2, and 3), and with the graph titles (item 11). Most importantly, participants had difficulty plotting a correct graph (item 8), and in the case of the Multiple-Baseline Graph task, they had difficulty determining and displaying the required conditions. These results suggest that the instruction needs to be revised to address these issues. The ideal response would be to optimize the current instruction as opposed to increasing the length of the instruction.

The rubric item averages revealed that participants were only moderately successful at the non-trivial aspects of graphing. This result was consistent with the feedback that suggested that the instruction was too long, and covered too many different aspects of graphing. It must be noted that participants could get high scores even if they plotted incorrect points on their graphs. Considering that the instructional content and the assessment rubric were both focused on the entire process of creating a graph, a participant could receive a high final score if they correctly constructed the axes and labels but plotted wrong points. The decision to make the weight of item 8 (correctly plotting points) the same as the other aspects of the graphing, seemed reasonable at the time. In a more realistic setting, the concept of what constitutes a correct graph would be different and in that scenario, the rubric would reflect the increased importance of plotting the correct points.

The overall average for the Skills Assessment was 67.71% (SD=15.46%). Chapter three established the criterion for effectiveness as a 10% increase from pretest to posttest, but because the Skills Assessment was not administered before the instruction, thus the criterion cannot be evaluated. The usefulness of a pretest for the Skills Assessment is questionable because even if the results were statistically and educationally significant, an aggregate posttest score of 67.71% still did not represent a passing grade.

There are many possible reasons for the poor performance on the Skills Assessment. The instruction, the assessment instrument, and the rubric are all areas that should be examined for potential problems. The first area of concern, however, was the length of the instruction. Participant feedback suggested that the single biggest issue was that the instruction was too long and this might have consequently resulted in sub-par performance.

Survey Development

There were significant differences in the proposed design of the survey and the actual design. Most of the changes were the result of the overall change in the design of the study. The original proposal involved a more systematic refinement process where items would be revised or eliminated in tandem with the development of the LOGIS application. In reality, the survey was created, refined, then implemented, thus the items did not change during the one-to-one evaluation, the small group evaluation, or the field trial.

As proposed, the survey comprised three parts: attitude towards graphing, attitude towards the interface, and attitude towards the LOGIS application. The original survey created by the researcher contained 14 items, 5 items focused on attitude towards graphing, 4 items focused on attitude towards the interface, and 5 items focused on attitude towards the LOGIS application. The 14 items were evaluated by nine students enrolled in a graduate level measurement class at the researcher's university. The students were given a paper copy of the survey items by their professor and they reacted to the items, making corrections, noting observations, and making suggestions on the paper containing the items. The students detected several typographical errors and noted that certain sentences had to be revised because they contained technical language that might be unfamiliar to the participants. They also highlighted other issues including tense shifting between items, items that were wordy, items that seemed to focus on ability and not attitude, and items that asked to participant to make a choice but did not provide alternatives from which to choose. Some students questioned the use of negative statements, but did not cite the negative statements as errors.

The revised items were then examined by an advanced doctoral student in the measurement and evaluation department at the university. The doctoral student commented on the revised items and confirmed some of the concerns raised by the previous nine reviewers. The doctoral graduate student recommended that the survey be extended to included items related to specific aspects of the interface and specific aspects of the application. The rationale was that participants would find it difficult to respond to some items if they, for example, liked the icons but disliked the buttons. The doctoral graduate student suggested that the survey would not be able to capture details regarding participant preferences because the survey did not contain enough items. Given the nature of the study and the time and resources needed to create and validate a larger survey, the researcher decided to forgo expanding the survey and instead concentrate on improving the current items.

The revised items were then examined by a member of a research consultancy group at the researcher's university. The consultant confirmed the doctoral student's concern regarding the number of items needed to fully address the interface and the application, but was comfortable with the survey given its purpose. The consultant identified some errors and made suggestions regarding the structure of certain items.

Based on the recommendations from the doctoral student and the research consultant, the survey was reduced to 11 items. Of the 11 survey items, 4 focused on attitude towards graphing, 3 items focused on attitude towards the interface, and 4 items focused on attitude towards the LOGIS application. Negatively worded items (3, 7, and 11) were kept because their advantages were considered to outweigh their disadvantages, but they were recoded during analysis to reflect a positive response. Table 16 shows the final survey items and their categories.

Table 16

The Items in the Final Version of the Survey

Survey Item	Category
1. It is important to know how to graph data. 2. Graphing is a useful skill to know. 3. I avoid graphing data whenever possible. 4. In the future I would create a graph to solve a problem when it is appropriate and useful	Attitude towards graphing
5. The interface is easy to use. 6. In the future I would use this interface if it were available. 7. The interface is difficult to use.	Attitude towards the interface
8. I think the application is useful. 9. This application helped me learn about graphing. 10. In the future I would use this application if it were available. 11. The application is unhelpful.	Attitude towards the application

After considering the suggestions of the nine graduate students, the researcher concluded that participants might not be familiar with the terms *interface* and *application*. The researcher decided to precede each item in those categories with a small descriptor. "The word interface refers to the text, buttons, icons, etc. that help you interact with a software program." preceded items in the attitude towards the interface category, and "The word application refers to the interface and the instructional content of a software program." preceded items in the attitude towards the application category.

Consistent with the proposal, the survey was delivered using a digital slide. Instead of the proposed 1.000 to 5.000 scale, the slide was scaled using integers from 1 to 1000 with incremental steps of 1 unit. There is a loss of precision when moving from 4000 discrete points (1.000 to 5.000) to 999 discrete points (1 to 1000), but the trade-off was acceptable because it was easier to program, manage, and interpret integers than decimals. It must be noted that the decision to use the 1 to 1000 scale was arbitrary.

The survey was originally designed to have one item with an unstructured response format but that feature was omitted because of programming difficulty. The ability to capture comments and free-form reactions would have been a good feature, but the same data were gathered during the one-to-one evaluations and the small-group evaluations.

The major validity threat to the instrument was the Location threat. All participants did not complete their survey in the computer lab, thus there is a possibility that their location influenced the results. This threat was not considered to be significant and no effort was made to minimize the potential effects.

The survey was delivered before and after the instruction, that is, participants experienced a pre-survey and a post-survey. The pre-survey contained the 4 items related to attitude towards graphing and the post-survey contained 11 items covering all three survey categories. Of the 47 participants who participated in the study, 45 successfully completed both the pre-survey and the post-survey. Each of the remaining 2 participants returned data that were missing one response, thus all their survey responses were excluded from the survey analysis.

The survey responses were based on an interval scale from 1 to 1000 with increments of 1 unit. This design is significantly different from the more common Likert scale design. Surveys that use the Likert scale return data that are ordinal in nature, but LOGIS was designed to return interval data. The researcher envisioned that using an interval-level scale would provide participants with greater control over their responses. The idea was that participants could fine-tune their responses and ensure that the survey slider was at the position that accurately reflected their attitude. The distinction is important because the level of measurement significantly influences how data are analyzed and subsequently interpreted. The decision to use an interval scale from Strongly Disagree (1) to Strongly Agree (1000) was based on the rationale that at minimum the data could be converted to the ordinal 5-point Likert scale by collapsing the data into 5 equal ranges. Where possible, both ordinal and interval analysis are included. To obtain ordinal data, the interval range was divided into five equal sections and then assigned a value based on the Likert scale. Table 17 shows the interval ranges that correspond to the Likert scale.

Table 17

The Likert Scale and Corresponding Interval Ranges

Likert Label	Likert Scale	Interval Range
Strongly Disagree	1	1 to 200
Disagree	2	201 to 400

Likert Label	Likert Scale	Interval Range
Neutral	3	401 to 600
Agree	4	601 to 800
Strongly Agree	5	801 to 1000

Descriptive statistics about data gathered from the post-survey responses are shown in Table 18.

Table 18

A Summary of the Post-Survey Data

Survey Item	*n*	*M*	*SD*	*Mdn*
1. It is important to know how to graph data.	45	798.87	236.44	856
2. Graphing is a useful skill to know.	45	815.64	211.39	848
3. I avoid graphing data whenever possible.	45	608.27	283.91	619
4. In the future I would create a graph to solve a problem when it is appropriate and useful	45	707.27	262.36	746
5. The interface is easy to use.	45	680.11	252.91	738
6. In the future I would use this interface if it were available.	45	699.51	271.89	750
7. The interface is difficult to use.	45	712.69	258.49	759
8. I think the application is useful.	45	721.04	266.49	793
9. This application helped me learn about graphing.	45	741.93	257.02	793

Survey Item	*n*	*M*	*SD*	*Mdn*
10. In the future I would use this application if it were available.	45	730.78	259.15	776
11. The application is unhelpful.	45	804.49	220.84	869

The post-survey data showed that in general, participants had high positive attitudes towards graphing, the interface and the application. The internal consistency reliability of the 11-item post-survey was.89. A Factor Analysis using Principal Components Analysis was conducted to determine if the 11 items were measuring the attitude categories and also to highlight items that need revision or items that should be eliminated. The sample size (n = 45) is somewhat small, but this procedure should provide an initial estimate of the reliability of the instrument. The attitude categories were treated independently but it is clear that there is a relationship between the application and the interface; the interface is included in the application. This initial analysis assumes that the categories are independent, hence the Varimax rotation, but it is expected that further revision of the items would lead to deeper critique of relationship between the application and the interface. The Kaiser-Meyer-Olkin (KMO) statistic and Bartlett's test revealed that the factor analysis was appropriate for these data. The KMO statistic was .763 for the data implying that reliable factors could be found from the data. The Bartlett's test of sphericity was highly significant ($p < .001$) for the data, thus there were valid relationships among the variables. Table 19 shows the rotated component matrices for the data excluding loadings that were less than .5.

Table 19

Rotated Component Matrices of the Post-Survey Data for the 45 participants

Survey Item	Component 1	2	3
1. It is important to know how to graph data.	.780		
2. Graphing is a useful skill to know.	.856		
3. I avoid graphing data whenever possible.	.834		
4. In the future I would create a graph to solve a problem when it is appropriate and useful	.856		
5. The interface is easy to use.		.818	
6. In the future I would use this interface if it were available.		.630	
7. The interface is difficult to use.		.896	
8. I think the application is useful.		.656	.590
9. This application helped me learn about graphing.			.934
10. In the future I would use this application if it were available.			.652
11. The application is unhelpful.			.823

The three components that were revealed corresponded to the three attitude categories. The Cronbach's alpha for the three scales were: attitude towards graphing (component 1) was .88, attitude towards the interface (component 2) was also .88, and attitude towards the application (component 3) was .86. Item 8 “I think the application is useful” was loaded on both attitude towards the interface and attitude towards the application, highlighting the need for closer examination of that item. Based on the data, the initial analysis of the post-survey suggested that the survey was consistent with its intended purpose.

As a point of interest, the researcher completed a factor analysis of the ordinal data; the interval data converted to Likert scales. The factor analysis revealed 2 components for the ordinal data. Component 1 (items 7 to 11) was a combination of attitude towards the interface and attitude towards the application, and component 2 (items 1 to 4) was attitude towards graphing. Component 1 had a Cronbach’s alpha of .89 and component 2 had an alpha of .85. The results of the ordinal data highlighted the possible relationship between the attitude towards the interface and the attitude towards the application. It is expected, however, that future evaluations will more clearly define and separate the interface construct from the application construct.

The pre-survey contained the first 4 items of the survey and focused on attitude towards graphing. These four items formed the basis from which the pre-survey and the post-survey were compared. Descriptive statistics about data gathered from the pre-survey responses are shown in

Table 20.

Table 20

A Summary of the Pre-Survey Data

Survey Item	*n*	*M*	*SD*	*Mdn*
1. It is important to know how to graph data.	45	778.87	227.74	788
2. Graphing is a useful skill to know.	45	798.31	233.76	826
3. I avoid graphing data whenever possible.	45	591.20	265.25	653
4. In the future I would create a graph to solve a problem when it is appropriate and useful	45	653.07	288.92	691

The internal consistency reliability (Cronbach's alpha) of the 4-item pre-survey was .84, suggesting the survey was consistent with its intended purpose. The pre-survey data showed that in general, participants had high positive attitudes towards graphing. This result suggested that there might not be a significant difference between the attitude towards graphing before the instruction and the attitude towards graphing after the instruction.

A comparison of the pre-survey and the first four items of the post-survey revealed that the responses were very consistent, that is, the scores were similarly distributed. Figure 20 shows a comparison of the pre and post survey responses across the first four survey items. The intervals on the abscissae correspond to the Likert scale intervals.

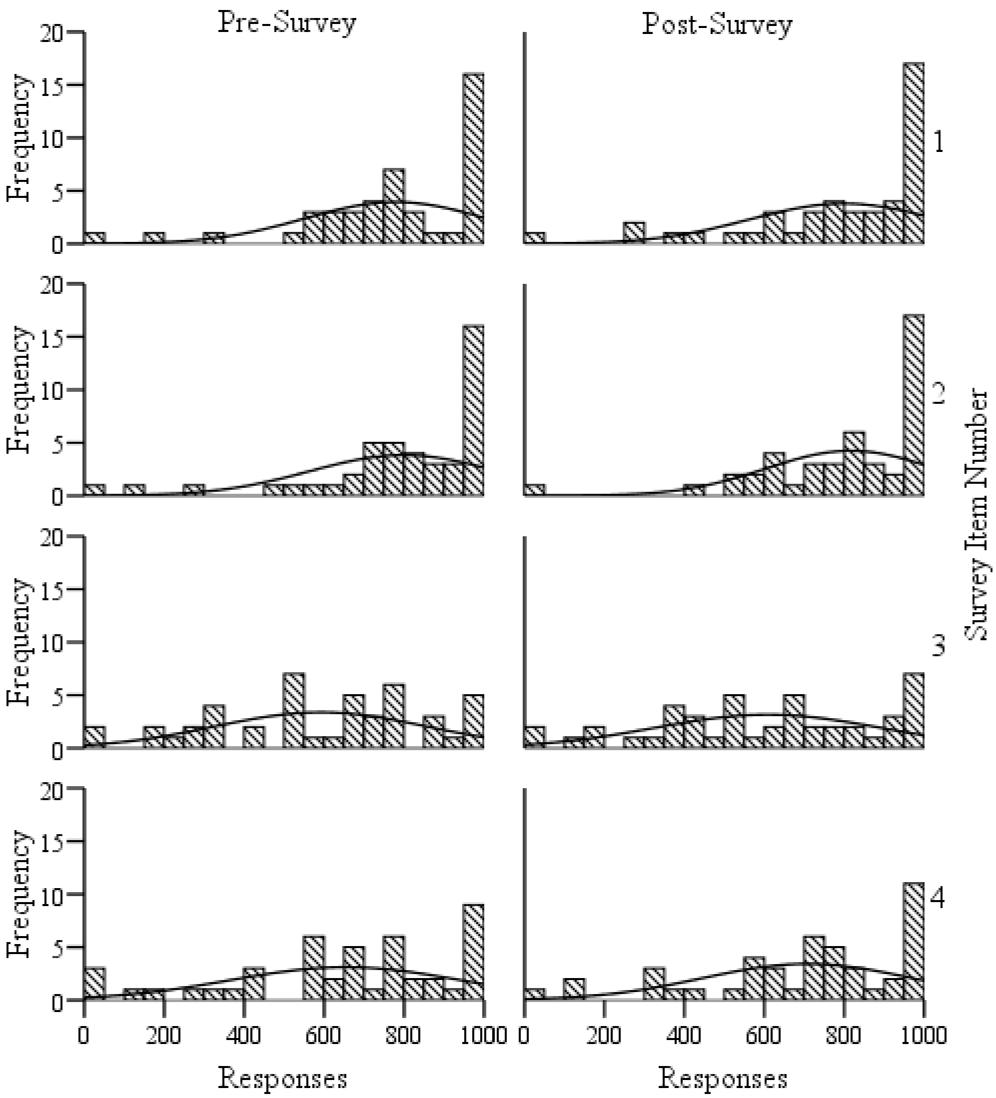


Figure 20. Histograms of the pre-survey and post-survey showing the distribution of the responses across the survey items ($n = 45$).

Figure 21 is a Scatterplot of the average pre-survey responses and average post-survey responses for each participant. The Scatterplot shows that the responses cluster in the area representing high attitude towards graphing before and after the instruction.

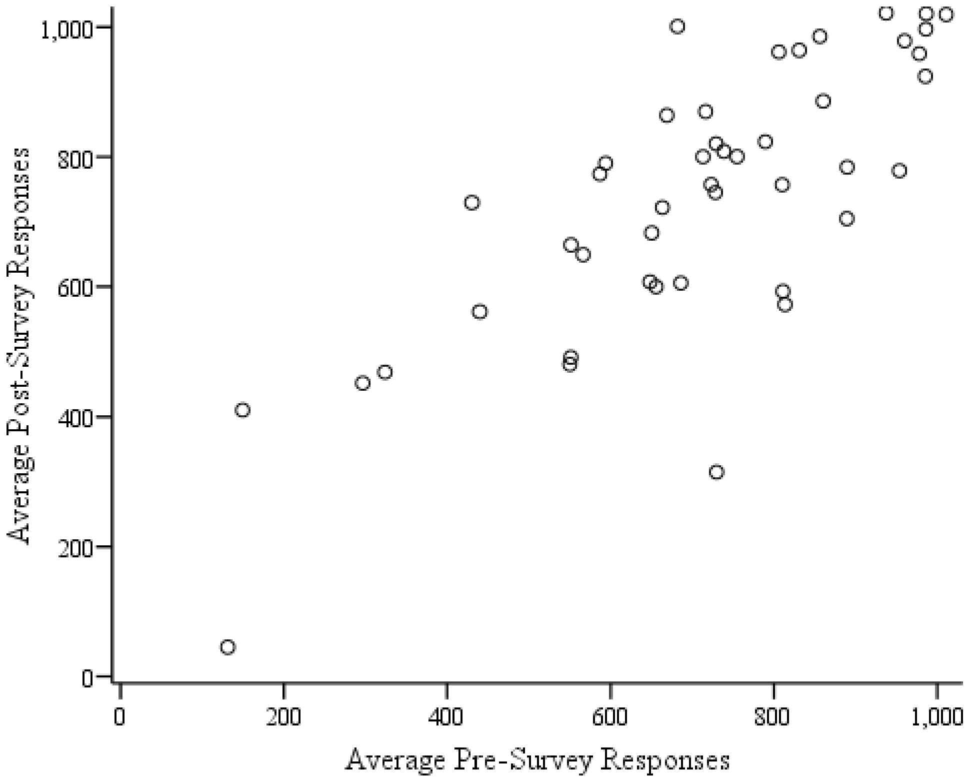


Figure 21. A Scatterplot of the average pre and post survey responses.

The pre-survey and post-survey were also examined in an effort to determine if there were significant differences in the responses. The data were analyzed using the paired-samples *t* test procedure. The *t* test revealed no significant differences between the pre-survey and post-survey responses on any of the survey items. The p-value for each survey item was greater than the .05 alpha level thus the failure to reject the null hypothesis that the means for the pre-survey and post-survey were different. Table 21 shows the results of the *t* test analysis.

Table 21

Results from the T Test Analysis on the First 4 Items on the Pre-Survey and Post-Survey

Survey Item	Mean Diff.	SD of Diff.	t	df	Sig. 2-tailed	Cohens *d*
1. It is important to know how to graph data.	-20.00	194.57	-0.69	44	.49	-.102
2. Graphing is a useful skill to know.	-17.33	178.95	-0.65	44	.52	-.097
3. I avoid graphing data whenever possible.	-17.07	224.49	-0.51	44	.61	-.076
4. In the future I would create a graph to solve a problem when it is appropriate and useful	-54.20	234.38	-1.55	44	.13	-.231

The data confirmed that with respect to the attitude towards graphing construct, the pre-survey and post survey responses were not different. The mean post-survey response (*M* = 732.51, *SD* = 213.86) was 3.85% greater than the mean pre-survey response (*M* = 705.36, *SD* = 210.31). The difference between the means is less than the 10% benchmark established in chapter three, thus the initial effectiveness of LOGIS in terms of attitude change could be viewed questionable. The researcher does not suggest that LOGIS is without value. This initial analysis reveals that the fundamental constructs of the survey and the survey items themselves need to be examined and refined.

Reflecting on the instruction, it is now clear that of the seven modules participants experienced, only one “The Importance Of Graphing” contained content that focused on the rationale for graphing. Of the 360 instructional frames that participants completed, only the 46 frames in the “The Importance Of Graphing” module addressed the issue of to attitude towards graphing. It is ambitious to think that an attitude could be changed when less than 13% of the instructional content was dedicated to that purpose.

Graphing Proficiency Test Development

The purpose of the Graphing Proficient Test was to determine the starting position of the participants. If a participant scored 100%, an arbitrarily chosen standard, on this test they would then be exempt from completing the Graphing Basics module. No participant scored 100%, therefore all the participants were required to completed the Graphing Basics module. There was no proposed development guide for the Graphing Proficiency Test, thus it was not systematically developed. This mistake on the part of the researcher significantly affected the value of this instrument. The final version of the test consisted of 10 items, 8 were Short-Answer items and 2 were Alternate-Choice items. The final Graphing Proficiency test items are listed in Appendix P.

Errors were found in the first version of the test and those items were revised upon discovery. The first 16 participants who completed the study experienced the first version of the proficiency test while the remaining 31 participants experienced the final version. Due to a programming syntax error, the last frame on the test returned erroneous data for 33 participants. It must be noted that the first and last frames of the Graphing Proficiency Test were not true test items. The first item was the introduction frame explaining the test, and the last item was the conclusion frame explaining the next step in the instruction for participants. The internal consistency reliability of the proficiency test, the Cronbach's alpha, was .46. Table 22 describes the overall performance on the second version of the Graphing proficiency Test.

Table 22

A Summary of the Statistical Data from the Graphing Proficiency Test

Source	*n*	Min	Max	*Mdn*	*M*	*SD*	skewness	kurtosis
Graphing Proficiency Test	47	20.00	90.00	50.00	52.77	18.38	0.36 SE = 0.35	-1.02 SE = 0.68

Figure 22 is a histogram of the grade distribution for the Graphing Proficiency Test showing the performance of the participants.

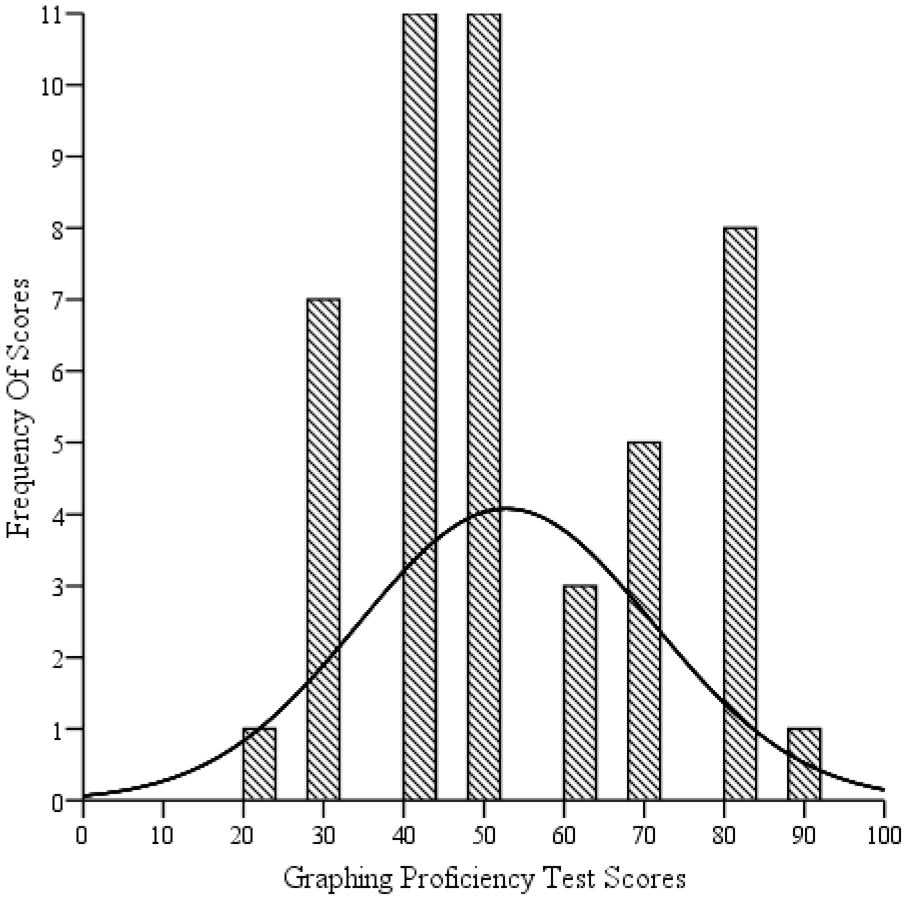


Figure 22. A Histogram of the proficiency test scores showing the distribution of the scores.

The lack of a development plan for the proficiency test negatively affected the resulting data. The internal consistency was low and participants performed poorly on the test. Based on the data, it is expected that this test would be significant revised.

Descriptive Information From The One-To-One Evaluation

Descriptive information was collected using questions introduced in Dick et al. (2005, p. 283):

1. Did you have any problems with specific words or phrases?
2. Did you have any problems with specific sentences?

3. Did you understand the themes presented in the instruction?
4. How effective was the sequencing of the instruction?
5. How would you describe the delivery pace of the instruction?
6. Was sufficient content presented in each section?
7. Was there sufficient time to complete the instruction?
8. What is your overall reaction to the instruction?
9. What are specific strengths of the instruction?
10. What are specific weaknesses of the instruction?

In addition to the reacting to preceding 10 questions on paper, the three participants in the one-to-one evaluation were asked to verbally comment on the interface, the content, the functionality of the application. They also reported their likes, dislikes, and changes they would make to the application. Two of the participants in the one-to-one evaluation had course averages of A at the time of the evaluation. The third participant had a course average of D at the time of the evaluation. The actual distribution of participants in the one-to-one evaluation was slightly different from the proposed distribution. The proposed distribution included one student that was a medium performer, but this was not possible due to scheduling conflicts. Appendix Q shows the three participants' actual written reactions to the 10 questions.

The one-to-one evaluation group provided significant help in getting the LOGIS to work properly. They identified, for example, content errors, grammatical errors, and typographical errors. Observing the participants in this group use the application revealed errors that affected program flow and syntax errors that affected program function. The contributions of the one-to-one group were significant, but most of the important revelations occurred as a direct result of observing the participants as they worked. For example, watching these early participants manipulate the mouse to complete the practice tasks led to the realization that the original 1-pixel fidelity setting was too exact. The first participant completed some of the graphing practice tasks while the fidelity setting was one pixel. This fidelity setting required that the participant locate individual pixels on the monitor. The researcher noticed that the participant was becoming increasingly frustrated because of the eyestrain that resulted from having to concentrate intensely on the computer monitor. The researcher changed the fidelity setting to 10-pixels allowing for a margin of error when graphing. The participant completed the instruction with a significantly reduced level of frustration. The identification and correction of this issue would not have been possible if the researcher did not directly observe the participant.

The importance of textual or verbal reactions cannot be minimized, but based on the textual responses in Appendix Q it is clear that observing the participants as they worked was a superior way of gathering data than written reactions by the participants. The original intent of the questions was to generate user feedback that could be used when revising the application, but it would be very hard to revise the application based on the written feedback in Appendix Q. The questions did not readily lend themselves to lengthy written responses thus participants' responses were very short and on occasions cryptic. It is apparent that the questions were too general for this application, and that more useful data could have been recovered if the questions were more specific to this application.

In addition to the questions, the participants were asked to verbally respond to 6 aspects of the task: the interface, the content, the functionality of application, their likes, their dislikes, and changes they would make. Appendix R is summary of the verbal reactions given by the one-to-one evaluation group.

The data obtained from the informal interview (see Appendix R) were significantly superior to the data obtained from the written questions. The participants addressed specific issues and made suggestions that they felt would make the application better. It is possible that the unstructured nature of the interview resulted in less inhibition on the part of the participants, or it is possible that the 6 aspects were sufficiently task-specific to generate useful feedback. Regardless of the reasons, the data obtained from the informal interview were extremely helpful in fixing immediate critical errors and also the data helped shaped an overarching view of the ways in which the application could be optimized.

The one-to-one evaluation is used to detect obvious errors in instruction and also to obtain initial performance data and user reactions. It is now clear that to achieve those objectives, questions requiring written responses need to be worded such they elicit rich responses from participants. The data from this evaluation suggested that in addition to the LOGIS application iteslf, the evaluation items also needed revision. Additionally, the participants must possess a basic knowledge about the application, the process, or both so that their critique is well informed.

Descriptive Information From The Small-Group Evaluation

Similar to the one-to-one evaluation group, the 13 participants who were in the small-group evaluation group wrote reactions to the 10 questions introduced in Dick et al. (2005, p. 283). Appendix S is a summary of the participants' responses. Similar to the one-to-one evaluation, most reactions were very short and sometimes cryptic. Some participants who provided detailed feedback but this seemed to be a function of the participant and not a result of the evaluation items.

The one-to-one evaluation and small-group evaluation provided useful data despite the shortcomings of the 10 questions. Apart from the typographical, content, and application errors that were revealed, participants overwhelmingly reported that the instruction was too long. In defense of the instruction, it must be noted that under normal circumstances participants would not be required to complete all the modules at one sitting, thus the length of the instruction would not be a significant issue. The most important issue revealed was the relationship between the number of frames and the number of modules. Participants preferred more modules and fewer frames per modules even if the total number of frames remained the same. This revelation should significantly influence future versions of LOGIS, as the instruction should now move towards more modules and fewer frames per modules.

Development Phase Critique

1. Was the rationale for including or omitting this phase clear?

The rationale for including this phase was clear. This phase could not have been omitted.

2. Are all required elements of this phase accounted for?

The required elements of the Development phase were all necessary and consequently they were all completed

3. Was this phase temporally cost effective to complete?

It was difficult to estimate the effectiveness of this phase because the researcher expected that this phase would consume most of the overall development time. The development phase took approximately 18 weeks and that translated to about 55% to the total time. More than 50% of the time in this phase was spent programming and testing the application. The remaining time was divided among developing the content, the assessments, and the surveys.

4. Are the products from this phase observable and measurable?

The Development phase resulted in goals that were observable and measurable. The application, content, and assessments were all artifacts of this phase.

5. Are the products from this phase consistent with the overall goal of the instructional application?

The products of the stage were consistent with the overall goal of the application.

Implement Phase Report

Gagne et al. (2005, p. 22) listed the two components of the Implement Phase: Market materials for adoption by teachers or students, and Provide help or support as needed. These two components were not addressed during the current study.

Implement Phase Critique

1. Was the rationale for including or omitting this phase clear?

This study did not go beyond the design and development of LOGIS, thus the Implement phase was not required.

2. Are all required elements of this phase accounted for?

   The question is not applicable.

3. Was this phase temporally cost effective to complete?

   The question is not applicable.

4. Are the products from this phase observable and measurable?

   The question is not applicable.

5. Are the products from this phase consistent with the overall goal of the instructional application?

   The question is not applicable.

Evaluate Phase Report

The actual implementation of the Evaluation Phase components occurred in the Development phase. This did not occur by design; no other solution seemed practical and reasonable. Considering that the products of the Evaluation phase were all used in the revision process, the role of a separate and independent Evaluation phase becomes unclear. It is important to note that the overarching question is not the necessity of evaluation rather the question is whether an evaluation phase is necessary. The case could be made that most of the evaluation should occur during the development phase. An evaluation would only be used if, for example, the application were being experimentally evaluated.

Evaluate Phase Critique

1. Was the rationale for including or omitting this phase clear?

The components of this phase were implemented in the Development Phase because no other solution could be found.

2. Are all required elements of this phase accounted for?

The “Implement plans for unit (course) maintenance and revision” component was not implemented because it was outside the scope of this study.

3. Was this phase temporally cost effective to complete?

The components of the phase were cost effective to complete.

4. Are the products from this phase observable and measurable?

The products from this phase were observations and measurements.

5. Are the products from this phase consistent with the overall goal of the instructional application?

The products of the stage are consistent with the overall goal of the application.

Summary

This chapter outlined the actual model-based development of LOGIS. The prescriptive nature of the documentation of phases provided an avenue to answer the research question and helped to organize the development process. The researcher viewed the creation of instruction as a dynamic and ongoing process that required constant evaluation and revision, thus the Evaluation phase was omitted and its components implemented in the Development phase.

The data from this initial analysis of LOGIS revealed that the application has the potential to be effective if it is revised accordingly. The data revealed a statistically and educationally significant increase in the Knowledge component although the mean posttest score was low. The Skills component could not be tested because no pre-measure was taken, but data revealed a low mean score on this component. The data from the Attitude component revealed no significant difference from the pre-survey to the post-survey, but the overall ratings of attitudes towards graphing, the interface, and the application were positive. LOGIS was not evaluated using an experimental design thus the increases cannot be attributed entirely to the application. These data were produced from the first iteration of testing on LOGIS and it is expected that future revisions would increase the value of the application.

CHAPTER FIVE

CONCLUSIONS

Chapter Map

This chapter answers the two main research questions, presents a general reflection, and discusses future research directions. The research questions are discussed in practical terms and within a theoretical context. The following map describes the organization of the chapter:

- Conclusions
    - Chapter map
    - Research question 1
        - Organization
        - Time
        - Perspective
    - Research question 2
    - General Reflection
    - Future direction
    - Summary

Research Question 1

How does the use of the ADDIE model influence the creation of the LOGIS instructional application?

The ADDIE model had a significant impact on the creation of LOGIS. The influence of the model can be categorized into three groups: Organization, Time, and Perspective. The effects of the model can be concisely stated as:

Organization

- The model forced a high level of planning to occur.
- The model dictated the sequence of tasks to be accomplished. This reduced frustration because there was never a question of what to do next.
- The model facilitated the definition of terminal states and made it easier to determine when a task had been completed and when it was time to move to the next item.
- The model caused the researcher to constantly monitor the high-level view (the “big picture”) and the low-level details, and the way each affects the other.

Time

- The model influenced the overall development time.

Perspective

- The model forced the researcher to revisit previously held assumptions about the design of instruction.
- The model facilitated reflection and diagnosis after the study was completed.

Organization

Planning and organization are features inherent in the ADDIE model and following the components in each phase is one way to increase the probability that the project will result in good products. In theory, the model should provide a framework that guides the development of an application while facilitating a certain level of creativity on the part of the developer. In practice, the researcher can confirm that the ADDIE model provided a structure that kept the development process on target.

The importance of planning was clearly visible during the development process. The data derived from planned elements were consistent and meaningful, but data derived from under-planned elements, for example the Graphing Proficiency test, were not reliable. In the case of the Graphing Proficiency test, no participant received 100% thus they all had to complete the Basic Graphing module. The consequences of this under-planned test were not severe but in general a developer should ensure that every event is as planned and as controlled as possible. This speaks to the interconnectedness of the development process. If any element of the process is not sufficiently planned, the effects will ripple throughout the entire development process reducing the usefulness and effectiveness of the final product.

One clear advantage of ADDIE is the facilitation of transition between components and phases. The research was never lost during the development process and this was primarily because of ADDIE’s built-in definitions of terminal products. Using the model required that the researcher define terminal states, thus the researcher always knew when a task was completed and when the next task could begin. This capability was important because it provided a mechanism to ensure that individual components received their fair share of time and resources. The ability to estimate time and resources early in the development process made the planning process easier. The identification of terminal states was also important because it was a source of motivation. The completion of each component marked the beginning of another, and this was a significant motivator because each completion meant that the researcher was one step closer to completing the project.

LOGIS contains many elements, and each element was designed to function within the broader context of the application. The ADDIE model helped to create a situation where the researcher could concentrate on individual low-level tasks while maintaining a high-level perspective (the “big picture”) on the project. The benefit was that during the planning and execution of a component, the researcher could easily change perspective to see how the individual component would affect the overall project. Maintaining a high-level perspective allowed the researcher to identify potential problems and take corrective measures early in the process. This ability was critical because the researcher was the sole developer and was responsible for every aspect of the project. It could be argued that the use of a model is the most important consideration when a single developer is creating an application. This is exemplified by the fact that the structure of the ADDIE model provided the researcher with a mechanism for forecasting, thus helping to avoid significant problems in the later stages of development.

The researcher categorized developers as less-skilled, moderately-skilled, and skilled, where experience significantly contributes to but does not determine skill level. This researcher can be described as a moderately-skilled developer. The planning and organization features that are inherent in the ADDIE model are very helpful to less-skilled and moderately skilled developers. Developers who do not have sufficient organizational or project-management skills can use the model as a structured environment to complete their task. In this scenario, the model masks or compensates for some of the deficiencies of the developer by providing a structured framework that can be used as a pathway for successful development. The model’s disciplined and methodical approach to development is beneficial to less-skilled developers, but it might also be a source of frustration for those developers. The developer might feel restrained in a structured environment and this might lead to a situation where components or phases are eliminated because they initially appear unimportant. It is critical that less-skilled developers understand that the structured and systematic nature of the model is the source of model’s usefulness. If a quality product is expected, the less-skilled developer must resist the urge to take shortcuts by eliminate components.

It may be the case that moderately-skilled and skilled developers do not need to use a model or that they can be successful despite skipping components. It is clear that skilled developers possess a level of experience that translates into faster and more accurate development but this does not necessarily mean that a model is not needed, or that they do not use models. It can be argued that skilled developers use models and do not skip components. Their expertise and experience makes it appear as though they skip components when in fact the components are simply being completed efficiently, internally, or both. A moderately-skilled developer may have some of the skills necessary for fast and accurate development. These developers must constantly and critically evaluate themselves to ensure that they operate within the bounds of their skill level. If they assume that they are at a higher level, these moderately-skilled developers will be susceptible to the same dangers that less-skilled developers face.

The skill level of the researcher was an important variable in the development process. This researcher can confirm that although there were issues regarding workload, the project would have been significantly more difficult without the model. The ADDIE model acted as a mask and compensated for the moderate experience of the researcher. The organizational and project-management skills of the researcher were enhanced by the model and it was easy to stay on target. The researcher maintained the belief that if the ADDIE steps were earnestly followed, the resulting product would be valuable. This "blind faith" was necessary because the researcher did not have prior experience developing this type of instructional application.

The importance of the role of the researcher in the development process cannot be understated. In this project, the researcher was the analyzer, the subject matter expert, the designer, and the programmer. In this case, the management and execution of these roles placed extraordinary burden on the researcher and the temptation to eliminate components was always present. The constant re-visualizing of the development process and the constant forecasting of possible problems were very frustration at times. This was probably not due to the presence of the model, but more an artifact of the researcher as the sole developer. It could be argued that many issues related to workload and frustration would not exist if the researcher was not the sole developer.

The relationship between the need for a model and the role of the developer is complex. The researcher was the sole developer in this project and consequently the researcher was extremely reliant on the structure of the ADDIE model. It is difficult to comment of what would have happened if the researcher was not the sole developer, or if another model, or no model, was used in the development process. The researcher, however, can make suggestions that are based on the experiences gained during this study. It is reasonable to suggest that as the developer assumes a greater role and therefore a greater workload in a project, the developer should be more inclined to use the framework that a model provides. At minimum, the framework would provide a mechanism to organize the workload and ultimately help to create a successful product.

Time

It became clear, upon reflection, that using ADDIE and the steps outlined in Gagne et al. (2005) was not the most temporally efficient way to develop LOGIS. A significant amount of time was spent adapting the requirements of the project to components in the ADDIE phases. This situation was evident during the Analysis and Design phases where goals and objectives were clarified. During these phases, the researcher used the ADDIE components to redefine pre-existing goals in an effort to make the goals fit the model's structure. This task was not wasted effort because those redefined goals were critical during later phases. The key issue is to what degree a model accommodates a given project. It is clear that some models are better suited for certain tasks, but it can be argued that the generalness of the ADDIE model makes it usable for all projects, but not optimally suited for any project. In this case, the ADDIE model forces the developer to spent time generating and refining the products of the early phases, at the cost of development time. This extra time is necessary because the developer must compensate for ADDIE's generalness by ensuring that each element of each component is well defined and developed. The task of detailing every aspect of a project is inherently a worthwhile task, but it might not be necessary if a model is optimally suited for a given project. The ADDIE model is not optimally suited for any task thus its success is dependent upon the systematic, structured, and detailed completion of every aspect of a project.

This study addresses key arguments made by Gordon and Zemke (2000) against ISD. The researcher can confirm the claim by Gordon and Zemke (2000) that the model is rigid, but this was expected given that the model is systematic and linear in its approach to development. The researcher does not agree, however, that using the model as prescribed produces bad results. Bad results are the natural consequences of bad ideas, poor planning, haphazard development or faulty implementation. If followed earnestly, without skipping steps, the model will compensate for some of the less-skilled developer’s deficiencies but it cannot transform a bad idea into a good idea. Instead of focusing on the product, a more valid criticism of the model might be its effect on the development process. The model might have an adverse effect on the development process under certain conditions, but criticisms of the model cannot be made without commenting on the role of the developer in the project, the skill level of the developer, and the nature of the task. Those variables seem to influence the development process significantly and thus they should be included in any critique of the ADDIE model.

Perspective

Using the ADDIE model as outlined in Gagne et al. (2005) resulted in an analysis of the learning process and the nature of instruction. Completing the components in each phase and determining what to include and where, was literally an exercise in shaping a value system. In essence, working with ADDIE caused the researcher to make value judgments regarding learning and instruction. For example, the researcher had to reconcile the idea of evaluation; its place in the development cycle and its role in instruction. In this study, the researcher considered evaluation to be an important component in the instruction process but it was not limited to the evaluation of learners. The researcher concluded that the application, which includes the interface elements and their interaction, was just as important as the content in terms of the instruction and learning processes. Using this viewpoint, equal emphasis was placed on both the development and refinement of the application and the development and refinement of the content and assessment instruments. Consequently, learners were evaluated in terms of content (knowledge and skills tests) and in terms of their interaction with the application (survey). It is important that developers recognize the importance of the application in the instruction and learning processes. It is not unreasonable to suggest that well-designed content and assessment instruments are less effective and reliable if their delivery systems are not equally well-designed. The key perspective is that the content, the assessment, and the application (the interface elements and their interactions) are all equally important in the evaluation process.

Using the ADDIE model influenced the researcher’s view on the general nature of instructional applications. The ADDIE mode emphasizes the use of formative evaluation in the development process. If an evaluation reveals a deficiency, the application must allow the developer to implement the required changes with relative ease, hence the concept of an Instructional Engine. LOGIS was designed to be flexible enough to accommodate a range of instruction. For example, if an evaluation revealed that learners were weak in the area of plotting points, a Module covering that topic could be easily added to LOGIS. This was the result of developing LOGIS such that it could accommodate instruction on any topic related to linear graphing. The underlying principle was the perspective that the application and the content could be decoupled such that the application could accommodate various content. The benefit of this perspective was that additional similar content could be easily added to LOGIS, and also the current content could be modified with full confidence that the changes would not adversely affect the application.

One artifact of using the model was the ability to critically look at the product and determine the location and cause of problems. The systematic nature of the model makes it easier to determine where errors occurred, the elements that contributed to the error, and the elements that were affected by the error. The survey data can be used to illustrate the point that the model made it easy to find the source of errors. If LOGIS was evaluated using an experimental design and the data revealed that there was no significant attitude change from pre-survey to post-survey, the researcher might be prompted to investigate how much instruction time was spent on attitude. Looking at the model, it would be easy to determine the exact point where adjustments should be made to cause a different outcome. In this case, an extended definition of the attitude goal would help emphasize the importance of attitude change, but more significantly, the Design phase component “Determine the instructional topics (or units) to be covered, and how much time will be spent on each” would be expanded to include more attitude based content.

Research Question 2

Is LOGIS an effective form of instruction?

a. Is LOGIS an effective form of instruction as measured by educationally significant differences in learners' performance from the pretest to the posttest on the Knowledge measure?

b. Is LOGIS an effective form of instruction as measured by educationally significant differences in learners' performance from the pretest to the posttest on the Skills measure?

c. Is LOGIS an effective form of instruction as measured by educationally significant differences in learners' attitude towards graphing from the pre to the post Graphing Attitude Survey?

The data from the analysis of LOGIS must be viewed within the correct context. The effectiveness of the LOGIS instructional application was determined with several consideration. Firstly, LOGIS was not experimentally evaluated thus these initial data results do not imply causality on the part of the application. The initial data results simply gauge the initial value of LOGIS and provide a basis for comment on the potential of the application. Secondly, under normal circumstances, participants would not need to complete the entire instruction in one sitting. It is possible that the time taken to complete the task affected the performance of the participants. Finally, this study represents the first iteration in the development of LOGIS. Further revision will increase the value of the application.

The data in this initial analysis of LOGIS revealed that the application was more effective and useful in terms of the Knowledge measure than in terms of the Skills and

Attitude measures. The Knowledge Assessment posttest scores (*M* = 62.41, *SD* = 11.96) were significantly higher than the pretest (*M* = 43.36, *SD* = 8.22). Although the participants performed statistically and educationally (a 10% increase from pre to post test) better on the posttest than they did on the pretest, their overall performance on the posttest was below average. These initial data revealed that several items should be revised or eliminated in an effort to continue the development of a valid and reliable knowledge assessment instrument.

There was no pretest for the Skills task thus the criteria for educational significance (a 10% increase from pre to post test) could not be assessed. Overall performance on the Skills Assessment (posttest) was average. Performance on the Cumulative Graph task was average (*M*=75.76%, *SD*=13.78), while performance on the Multiple-Baseline Graph task was below average (*M*=61.21, *SD*=29.35). These initial data suggest that the tutorial content and the grading rubric should be examined and refined to increase the value of the LOGIS.

Participants had positive attitudes towards graphing, towards the interface, and towards the application. The post-survey responses were 3.85% greater than the pre-survey responses thus the increase was not educationally significant.

General Reflection

The current study used the ADDIE components outlined by Gagne et al. (2005) verbatim. Under normal circumstances, a developer would probably adjust the model's components to meet the needs of the task, resulting in at least a reduction in frustration level. The resulting development process would reflect the developer's interpretation of model possibly resulting in an optimal solution to the task. The researcher's use of the ADDIE components verbatim is a strong testimony to the fact that the model is flexible enough to produce good results even when a very constrained interpretation of its components is applied. This suggests that the model increases in value when it is optimized for a particular task.

The effect of the model on the development process was categorized by Organization, Time, and Perspective. In addition to the categories, the researcher identified three variables that also contribute to the relationship between the developer and the model: the researcher as the sole developer, the skill level of the developer, and the nature of the task. The relationships among the three variables, the model, and the developer are unclear but the researcher contends that they are important considerations.

In terms of the researcher as the sole developer, it is reasonable to suggest that a team implementation of ADDIE would have significant benefits in terms of the development process. Workload and subsequent frustration levels are immediately reduced when the team approach is implemented. In addition, each team member can focus on specific tasks resulting in optimal solutions for each aspect of the project. The most significant drawback of the team approach is the issue of organization. If organization and communication issues exist, the development process is more prone to, for example, delays caused by duplicated effort. In the team approach, organization and communication should be at the forefront of development.

The effect of the researcher's skill-level on the development process is an important consideration. The researcher had extensive programming experience and it is unclear how this study could have been completed without significant programming ability. In this study, ADDIE's facilitation of forecasting was paired with the researcher's ability to critical and accurately assess future issues, resulting in a complete view of future factors that might affect the project. This was important because the researcher could predict problem areas and take action early to avoid them. The individual contributions of the researcher's abilities and the model's characteristics are unknown but it is clear that they both influence the development process.

One of the most important reasons for using is model is the consistency that it affords. Using a model should result in products that are optimal and reproducible. One important consideration is the degree to which the model minimizes the effects of the developer such that similar results can be obtained by different developers. Based on the results of this study, it can be stated that the use of the ADDIE model does not completely mitigate the effects of the developer, but does provide the mechanism through which different developers can produce similar results. This consideration can be extended to Developmental Inquiry in general. It is worthwhile to ask if a different developer followed the same process as this researcher, would the resulting product be similar to LOGIS? The idea of minimizing variance between researchers is pertinent to both the ADDIE model and Developmental Research. Unlike experimental methods, the ADDIE model and Developmental Research methods cannot eliminate the variance between researchers. They do, however, provide a level of specificity greater than that of extremely qualitative methods, hence different developers can produce results that are comparable.

The researcher reviewed literature that called for an increased emphasis on Developmental Research. This researcher agrees with the assessment that the

Instructional Technology field should place more focus on Developmental Research. From a practical perspective, the researcher did not have any significant frame of reference for visualizing or executing this study. Theoretical assertions are important and they were used in this study, but there was a lack of practical usable models that the researcher could use to steer the development process and the documentation of the process. This study could have benefited from an established body of scholarship because, for example, the importance of monitoring "time" as a variable was not an initial consideration. A more developed body of research would have prompted the researcher to include or exclude certain elements thereby increasing the value of the study.

This study can serve as one of the initial data-based examination of the development process. The hope is that it provides an additional perspective from which the development process can be examined.

Future Direction

Future research on this topic should include LOGIS, variables affecting ADDIE, and Developmental Research.

The LOGIS application will be revised based on the data and conclusions in this study. The interface will be retooled with more attention focused on font sizes, scrolling, and task instructions. The content will be revised to include more examples and to reflect better priming and prompting. The content will also be revised to more modules with fewer frames per module.

The assessments will be revised to increase their reliability. Knowledge assessment items that were bad will be revised or removed, and the total number of items in the assessment will be reduced. The Skills assessment rubric will be refined to reflect better standards and the grading scheme will be revised to reflect the increased importance of accurately plotting data points. The survey will be revised such that there is a clear distinction between the attitude towards the interface and attitude towards the application constructs. The Graphing Proficiency test will be revised to include more systematically developed items. Addressing these issues should provide a good foundation for the second iteration of LOGIS refinement.

Three variables appear to be significant in the development process: the researcher as the sole developer, the skill level of the developer, and the nature of the task. The relationship between these three variables and the use of a model should be examined carefully. Future researcher should try to determine the contributions of each variable to the development process.

Future research should include a more in-depth look at Developmental Research and its role within the field. More data-based research should be attempted in an effort the determine the true value and usefulness of Developmental Research

REFERENCES

Ajzen, I. (2001). Nature and operation of attitudes. *Annual Review of Psychology, 52*, 27-58.

Alberto, P. A., & Troutman, A. C. (2006). *Applied behavior analysis for teachers* (7th ed.). Upper Saddle River, New Jersey: Pearson Merrill Prentice Hall.

Alessi, S. M., & Trollip, S. R. (2001). *Multimedia for learning: Methods and development* (3rd ed.). Needham Heights, Massachusetts: Allyn and Bacon.

Barton, R. (1997). Computer-Aided Graphing: a comparative study. *Journal of Information Technology for Teacher Education, 6*(1), 59-72.

Bloom, B. S. (1956). *Taxonomy of educational objectives; the classification of educational goals* (1st ed.). New York,: Longmans Green.

Brasseur, L. (1999). *The role of experience and culture in computer graphing and graph interpretive processes.* Paper presented at the Proceedings of the 17th annual international conference on Computer documentation.

Brasseur, L. (2001). Critiquing the Culture of Computer Graphing Practices. *Journal of Technical Writing and Communication, 31*(1), 27-39.

Burton, J. K., Moore, D. M., & Magliaro, S. G. (2004). Behaviorism and instructional technology. In D. H. Jonassen (Ed.), *Handbook of research for educational communications and technology* (2nd ed., pp. 3-36). Mahwah, N.J.: Lawrence Erlbaum.

Campbell, J. O., Bourne, J., Mosterman, P. J., & Brodersen, A. J. (2002). The effectiveness of learning simulations for electronic laboratories. *Journal of Engineering Education, 91*(1), 81-87.

Clark, R. C. (2002). The new ISD: Applying cognitive strategies to instructional design. *Performance Improvement, 41*(7), 8-15.

Clark, R. E. (1983). Reconsidering Research on Learning from Media. *Review of Educational Research, 53*(4), 445-459.

Clark, R. E. (2005). Guided experiential learning: Training design and evaluation: University of Southern California, Institute for Creative Technologies.

Couture, M. (2004). Realism in the Design Process and Credibility of a Simulation-Based Virtual Laboratory. *Journal of Computer Assisted Learning, 20*(1), 40-49.

Crano, W. D., & Prislin, R. (2006). Attitudes and persuasion. *Annual Review of Psychology, 57*, 345-374.

Crocker, L. M., & Algina, J. (1986). *Introduction to classical and modern test theory*. New York: Holt Rinehart and Winston.

Cuban, L. (1986). *Teachers and machines: The classroom use of technology since 1920*. New York, New York: Teachers College Press.

Davis, D. R., Bostow, D. E., & Heimisson, G. T. (2005). *Experimental Evaluation of Incremental Prompting as a Feature of Web-Delivered Programmed Instruction (Applied Behavior Analysis).* Paper presented at the Association for Behavior Analysis, Chicago, Illinois.

Davis, D. R., Bostow, D. E., & Heimisson, G. T. (in press). Strengthening scientific verbal behavior: An experimental comparison of progressively prompted, unprompted programmed instruction, and prose. *Journal of Applied Behavior Analysis*.

de Jong, T., & van Joolingen, W. R. (1998). Scientific Discovery Learning with Computer Simulations of Conceptual Domains. *Review of Educational Research, 68*(2), 179-201.

De Vaney, A., & Butler, R. P. (1996). Voices of the Founders: Early Discourses in Educational Technology. In D. H. Jonassen (Ed.), Handbook of research for educational communications and technology : a project of the association for educational communications and technology (pp. 3-43). New York: Macmillan Library Reference USA.

Dick, W., Carey, L., & Carey, J. O. (2005). *The systematic design of instruction* (6th ed.). Boston, Massachusetts: Pearson/Allyn and Bacon.

Emurian, H. H., & Durham, A. G. (2003). Computer-based tutoring systems: a behavioral approach. In J. A. Jacko & A. Sears (Eds.), *The human-computer interaction handbook: fundamentals, evolving technologies and emerging applications* (pp. 677-697): Lawrence Erlbaum Associates, Inc.

Ericsson, K. A., Krampe, R. T., & Tesch-Römer, C. (1993). The Role of Deliberate Practice in the Acquisition of Expert Performance. *Psychological Review, 100*(3), 363-406.

Fraenkel, J. R., & Wallen, N. E. (2006). *How to design and evaluate research in education* (6th ed.). New York: McGraw-Hill.

Gagne, R. M., Wager, W. W., Golas, K. C., & Keller, J. M. (2005). *Principles of instructional design* (5th ed.). Belmont, California: Wadsworth/Thomson Learning.

Gal, I. (2002). Adult statistical literacy: Meanings, components, responsibilities. *International Statistical Review, 70*(1), 1-25.

Gilbert, J. E., & Han, C. Y. (1999). Adapting instruction in search of ‘a significant difference’. *Journal of Network and Computer Applications, 22*(3), 149-160.

Gordon, J., & Zemke, R. (2000). The Attack on ISD. *Training, 37*(4), 42-45.

Gredler, M. E. (2004). Games and simulations and their relationships to learning. In D. H. Jonassen (Ed.), *Handbook of research for educational communications and technology* (2nd ed., pp. 571-581). Mahwah, N.J.: Lawrence Erlbaum.

Hansen, R. E. (2000). The Role of Experience in Learning: Giving Meaning and Authenticity to the Learning Process in Schools. *Journal of Technology Education, 11*(2), 23-32.

Hartley, J. (1998). *Learning and studying : a research perspective*. London ; New York: Routledge.

Jarvis, P. (1987). *Adult learning in the social context*. London ; New York: Croom Helm.

Jarvis, P., Holford, J., & Griffin, C. (2003). *The theory & practice of learning* (2nd ed.). London ; Sterling, VA: Kogan Page.

Jones, T. S., & Richey, R. C. (2000). Rapid prototyping methodology in action: A developmental study. *Educational Technology Research and Development, 48*(2), 63-80.

Kirschner, P. A., Sweller, J., & Clark, R. E. (2006). Why minimally guided instruction does not work. *Educational Psychologist, 41*(2), 75-86.

Kozma, R. B. (1991). Learning with Media. *Review of Educational Research, 61*(2), 179-211.

Kritch, K. M., & Bostow, D. E. (1998). Degree of constructed-response interaction in computer-based programmed instruction. *Journal of Applied Behavior Analysis, 31*(3), 387-398.

Lane, D. M., & Tang, Z. (2000). Effectiveness of Simulation Training on Transfer of Statistical Concepts. *Journal of Educational Computing Research, 22*(4), 383-396.

Lee, J. (1999). Effectiveness of Computer-Based Instructional Simulation: A Meta Analysis. *International Journal of Instructional Media, 26*(1), 71-85.

Lockee, B., Moore, D., & Burton, J. (2004). Foundations of programmed instruction. In D. H. Jonassen (Ed.), *Handbook of research for educational communications and technology* (2nd ed., pp. 1099-1130). Mahwah, N.J.: Lawrence Erlbaum.

McDonald, J. K., Yanchar, S. C., & Osguthorpe, R. T. (2005). Learning from programmed instruction: Examining implications for modern instructional technology. *Educational Technology Research and Development, 53*(2), 84-98.

McGraw, K. O., & Wong, S. P. (1996). Forming inferences about some intraclass correlation coefficients. *Psychological Methods, 1*(1), 30-46.

McKenney, S., & van der Akker, J. (2005). Computer-based support for curriculum designers: A case of developmental research. *Educational Technology Research and Development, 53*(2), 41-66.

Merriam, S. B., & Caffarella, R. S. (1999). *Learning in adulthood : a comprehensive guide* (2nd ed.). San Francisco: Jossey-Bass Publishers.

Miller, D. W. (1999). The Black Hole of Education Research. *Chronicle of Higher Education, 45*(48), A17-A18.

Miller, M. L., & Malott, R. W. (1997). The importance of overt responding in programmed instruction even with added incentives for learning. *Journal of Behavioral Education, 7*(4), 497-503.

Mills, J. D. (2004). Learning Abstract Statistics Concepts Using Simulation. *Educational Research Quarterly, 28*(4), 18-33.

Molenda, M. (2003). In Search of the Elusive ADDIE Model. *Performance Improvement, v42 n5 May-Jun 2003.*

Monteiro, C., & Ainley, J. (2002). Exploring Critical Sense in Graphing. *Proceedings of the British Society for Research into Learning Mathematics, 22*(3), 61-66.

Morrison, G. R., Ross, S. M., & Kemp, J. E. (2004). *Designing effective instruction* (4th ed.). Hoboken, New Jersey: John Wiley & Sons, Inc.

Papanikolaou, K. A., Grigoriadou, M., Kornilakis, H., & Magoulas, G. D. (2003). Personalizing the Interaction in a Web-based Educational Hypermedia System: the case of INSPIRE. *User Modeling and User-Adapted Interaction, 13*(3), 213-267.

Park, O.-c., & Lee, J. (2004). Adaptive instructional systems. In D. H. Jonassen (Ed.), *Handbook of research for educational communications and technology* (2nd ed., pp. 651-684). Mahwah, N.J.: Lawrence Erlbaum.

Pimentel, J. R. (1999). Design of net-learning systems based on experiential learning. *Journal of asynchronous learning networks, 3*(2), 64-90.

Ray, R. D. (1995a). A behavioral systems approach to adaptive computerized instructional design. *Behavior Research Methods, Instruments, & Computers, 27*(2), 293-296.

Ray, R. D. (1995b). MediaMatrix: An authoring system for adaptive hypermedia teaching-learning resource libraries. *Journal of Computing in Higher Education, 7*(1), 44-68.

Reeves, T. C. (1995). *Questioning the Questions of Instructional Technology Research.* Paper presented at the Proceedings of the 1995 Annual National Convention of the Association for Educational Communications and Technology AECT), Anaheim, Ca.

Reeves, T. C. (2000a). *Enhancing the worth of instructional technology research through "design experiments" and other development research strategies.* Paper presented at the Symposium on International Perspectives on Instructional Technology Research for the 21st Century., New Orleans, LA.

Reeves, T. C. (2000b). Socially Responsible Educational Technology Research. *Educational Technology, 40*(6), 19-28.

Reeves, T. C. (2003). Storm Clouds on the Digital Education Horizon. *Journal of Computing in Higher Education, 15*(1), 3-26.

Reiber, L. P. (2004). Microworlds. In D. H. Jonassen (Ed.), *Handbook of research for educational communications and technology* (2nd ed., pp. 583-603). Mahwah, N.J.: Lawrence Erlbaum.

Reigeluth, C. M., & Frick, T. W. (1999). Formative research: A Methodology for Creating and Improving Design Theories. In C. M. Reigeluth (Ed.), *Instructional-Design Theories and Models: A New Paradigm of Instructional Theory* (Vol. 2, pp. 633-651). Mahwah, N.J.: Lawrence Erlbaum Associates.

Reynolds, C. R., Livingston, R. B., & Willson, V. (2006). *Measurement and assessment in education*. Boston, MA: Allyn & Bacon.

Richey, R. C., Klein, J. D., & Nelson, W. A. (2004). Developmental research: Studies of instructional design and development. In D. H. Jonassen (Ed.), *Handbook of research for educational communications and technology* (2nd ed., pp. 1099-1130). Mahwah, N.J.: Lawrence Erlbaum.

Roth, W.-M., & McGinn, M. K. (1997). Graphing: Cognitive Ability or Practice? *Science Education, 81*(1), 91-106.

Schunk, D. H. (2000). *Learning theories : an educational perspective* (3rd ed.). Upper Saddle River, N.J.: Merrill/Prentice-Hall.

Seels, B., & Richey, R. (1994). *Instructional technology : The definition and domains of the field*. Washington, D.C.: Association for Educational Communications and Technology.

Shute, V. J., & Psotka, J. (1996). Intelligent tutoring systems: past, present and future. In D. H. Jonassen (Ed.), *Handbook of research for educational communications and technology : a project of the association for educational communications and technology* (pp. 570-600). New York: Macmillan Library Reference USA.

Skinner, B. F. (1954). The science of learning and the art of teaching. *Harvard educational review, 24*, 86-97.

Skinner, B. F. (1957). *Verbal behavior*. New York,: Appleton-Century-Crofts.

Skinner, B. F. (1958). Teaching Machines. *Science, 128*, 969-977.

Skinner, B. F. (1972). *Beyond freedom and dignity*. Toronto ; New York: Bantam Books.

Skinner, B. F. (1986). Programmed instruction revisited. *Phi Delta Kappan, 68*(2), 103-110.

Smith, L. D., Best, L. A., Stubbs, D. A., Johnston, J., & Archibald, A. B. (2000). Scientific Graphs and the Hierarchy of the Sciences. *Social Studies of Science, 30*(1), 73-94.

Smith, M. K. (1999). Learning theory. Retrieved October 10, 2005, from http://www.infed.org/biblio/b-learn.htm

Smith, P. L., & Tillman, R. J. (2005). *Instructional design* (5th ed.). Hoboken, New Jersey: John Wiley & Sons, Inc.

Song, S. H., & Keller, J. M. (2001). Effectiveness of motivationally adaptive computer-assisted instruction on the dynamic aspects of motivation. *Educational Technology Research and Development, 49*(2), 5-22.

Spelman, M. D. (2002). GLOBECORP: Simulation versus Tradition. *Simulation & Gaming, 33*(3), 376-394.

Steinberg, R. N. (2000). Computers in teaching science: To simulate or not to simulate? *American Journal of Physics, 68*(S1), S37-S41.

Stemler, S. E. (2004). A comparison of consensus, consistency, and measurement approaches to estimating interrater reliability, *Practical Assessment, Research & Evaluation* (Vol. 9).

Stoyanov, S., & Kirschner, P. (2004). Expert Concept Mapping Method for Defining the Characteristics of Adaptive E-Learning: ALFANET Project Case. *Educational Technology Research and Development, 52*(2), 41-56.

Tomporowski, P. D. (2003). *The psychology of skill : a life-span approach*. Westport, Conn.: Praeger.

Trochim, W. M. (2002, January 16, 2005). Research Methods Knowledge Base. Retrieved March 6, 2006, from http://www.socialresearchmethods.net/kb/

Tudor, R. M. (1995). Isolating the Effects of Active Responding in Computer-Based Instruction. *Journal of Applied Behavior Analysis, 28*(3), 343-344.

Tudor, R. M., & Bostow, D. E. (1991). Computer-Programmed Instruction: The Relation of Required Interaction to Practical Application. *Journal of Applied Behavior Analysis, 24*(2), 361-368.

van den Akker, J. (1999). Principles and methods of development research. In J. van den Akker, R. M. Branch, K. Gustafson, N. Nieveen & T. Plomp (Eds.), *Design approaches and tools in education and training* (pp. 1-14). Dordrecht ; Boston: Kluwer Academic Publishers.

Wiesner, T. F., & Lan, W. (2004). Comparison of student learning in physical and simulated unit operations experiments. *Journal of Engineering Education, 93*(3), 195-204.

Wilson, B. G. (2005). Foundations for instructional design: Reclaiming the conversation. In J. M. Spector, C. Ohrazda, A. Van Schaack & D. A. Wiley (Eds.), *Innovations in instructional technology: Essays in honor of M. David Merrill* (pp. 237-252). Mahwah, New Jersey: Lawrence Erlbaum Associates.

Zemke, R., & Rossett, A. (2002). A Hard Look at ISD. *Training, 39*(2), 26-28,30-32,34.

Appendices

Appendix A: The Tutorials and Major Goals

- The Control And Measurement Of Behavior
    - Experimentation
    - Measurement is an important aspect of the science of behavior
    - Controlling behavior involves the manipulation of environmental variables
    - It is important to remain in contact with behavior
- The Importance Of Graphing
    - Data and graphing
    - The advantages of visual displays of data
    - statistical procedures versus visual presentations
    - Feedback and its importance
    - Variations in data
    - Interpretation of data
- Basic Graphing
    - Axes
    - Coordinates
    - Point
    - The origin
    - Scale of axes
    - Hatch marks
    - Slope of a line
    - The title and legend
    - Graphing data

Appendix A: (Continued)

- Behavioral Graphing Concepts
    - Data scale and path
    - Scale breaks
    - Dependent and independent variables
    - Functional relationship
    - Trends
    - The importance of time as a unit of measure
- The Cumulative Graph
    - The cumulative record
    - Upward slope of the depth
    - Difficulty in reading
    - Rate of responding and its effect on the graph
- The Multiple Baseline Graph
    - Graphing multiple datasets
    - Starting at the origin
    - Phases and their indications
    - The indication of special events

Appendix B: Guidelines for the Development of the Alternate-Choice Items

1. Avoid including more than one idea in the statement.
2. Avoid specific determiners and qualifiers that might serve as cues to the answer.
3. Ensure that true and false statements are of approximately the same length.
4. Avoid negative statements
5. Avoid long and/or complex statements.
6. Include an approximately equal number of true and false statements.
7. Avoid including the exact wording from the textbook

Appendix C: Checklist for the Development of the Alternate-Choice Items

1. Does each statement include only one idea? ________
2. Have you avoided using specific determiners and qualifiers that could serve as cues to the answer? ________
3. Are true and false statements of approximately the same length? ________
4. Have you avoided negative statements? ________
5. Have you avoided long and complex statements? ________
6. Is there an approximately equal number of true and false statements? ________
7. Have you avoided using the exact wording from the textbook? ________

Appendix D: Guidelines for the Development of the Multiple-Choice Items

1. Use a printed format that makes the item as clear as possible.
2. Have the item stem contain all the information necessary to understand the problem or question.
3. Provide between three and five alternatives.
4. Keep the alternatives brief and arrange them in an order that promotes efficient scanning.
5. Avoid negatively stated stems in most situations.
6. Make sure only one alternative is correct or represents the best answer.
7. Avoid cues that inadvertently identify the correct answer.
8. Make sure all alternatives are grammatically correct relative to the stem.
9. Make sure no item reveals the answer to another item.
10. Have all distracters appear plausible.
11. Use alternative positions in a random manner for the correct answer.
12. Minimize the use of "none of the above" and avoid using "all of the above."
13. Avoid artificially inflating the reading level.
14. Limit the use of always and never in the alternatives.
15. Avoid using the exact phrasing from the text.
16. Organize the test in a logical manner.
17. Give careful consideration to the number of items on your test.
18. Be flexible when applying these guidelines

Appendix E: The Print Format Guidelines

1. Provide brief but clear directions. Directions should include how the selected alternative should be marked.
2. The item stem should be numbered for easy identification, while the alternatives are indented and identified with letters.
3. Either capital or lowercase letters followed by a period or parenthesis can be used for the alternatives. If a scoring sheet is used, make the alternative letters on the scoring sheet and the test as similar as possible.
4. There is no need to capitalize the beginning of alternatives unless they begin with a proper name.
5. When the item stem is a complete sentence, there should not be a period at the end of the alternatives.
6. When the stem is in the form of an incomplete statement with the missing phrase at the end on the sentence, alternatives should end with a period.
7. Keep the alternatives in a vertical list instead of placing them side by side because it is easier for students to scan a vertical list quickly.
8. Use correct grammar and formal language structure in writing items.
9. All items should be written so that the entire question appears on one page.

Appendix F: Checklist for the Development of the Multiple-Choice Items

1. Are the items clear and easy to read? ________
2. Does the item stem clearly state the problem or question? ________
3. Are there between three and five alternatives? ________
4. Are the alternatives brief and arranged in an order that promotes efficient scanning? ________
5. Have you avoided negatively stated stems? ________
6. Is there only one alternative that is correct or represents the best answer? ________
7. Have you checked for cues that accidentally identify the correct answer? ________
8. Are all alternatives grammatically correct relative to the ________ stem?
9. Have you checked to make sure no item reveals the answer to another item? ________
10. Do all distracters appear plausible? ________
11. Did you use alternative positions in a random manner for the correct answer? ________
12. Did you minimize the use of "none of the above" and avoid using "all of the above"? ________
13. Is the reading level appropriate? ________
14. Did you limit the use of always and never in the alternatives? ________

Appendix F: (Continued)

15. Did you avoid using the exact phrasing from the text? ________

16. Is the test organized in a logical manner? ________

17. Can the test be completed in the allotted time period? ________

Appendix G: Guidelines for the Development of the Short-Answer Items

1. Structure the item so that the response is as short as possible.
2. Make sure there is only one correct response.
3. Use the direct-question format in preference to the incomplete-sentence format.
4. Have only one blank space when using the incomplete-sentence format, preferably near the end of the sentence.
5. Avoid unintentional cues to the answer.
6. Make sure the blanks provide adequate space for the student's response.
7. Indicate the degree of precision expected in questions requiring quantitative answers.
8. Avoid lifting sentences directly out of the textbook and converting them into short-answer items.
9. Create a scoring rubric and consistently apply it.

Appendix H: Checklist for the Development of the Short-Answer Items

1. Does the item require a short response? _________
2. Is there only one correct response? _________
3. Did you use an incomplete sentence only when there was no loss of clarity relative to a direct question? _________
4. Do incomplete sentences contain only one blank? _________
5. Are blanks in incomplete sentence near the end of the sentence? _________
6. Have you carefully checked for unintentional cues to the answer? _________
7. Do the blanks provide adequate space for the answers? _________
8. Did you indicate the degree of precision required for quantitative answers? _________
9. Did you avoid lifting sentences directly from the textbook? _________
10. Have you created a scoring rubric for each item? _________

Appendix I: Checklist for the Development of the Skills Assessment Items

1. Are the statements or questions an accurate representation? ________
2. Is the item appropriate and relevant to test specifications? ________
3. Are there technical item-construction flaws? ________
4. Did you use correct grammar? ________
5. Did you use offensive or bias language? ________
6. Is the level of readability appropriate? ________

Appendix J: Guidelines for the Development of the Survey Items

1. Put statements or questions in the present tense.
2. Do not use statements that are factual or capable of being interpreted as factual.
3. Avoid statements that can have more than one interpretation.
4. Avoid statements that are likely to be endorsed by almost everyone or almost no one.
5. Try to have an almost equal number of statements expressing positive and negative feelings.
6. Statements should be short, rarely exceeding 20 words.
7. Each statement should be a proper grammatical sentence.
8. Statements containing universals such as all, always, none, and never often intro-duce ambiguity and should be avoided.
9. Avoid use of indefinite qualifiers such as only, just, merely, many, few, or seldom.
10. Whenever possible, statements should be in simple sentences rather than complex or compound sentences. Avoid statements that contain "if" or "because" clauses.
11. Use vocabulary that can be understood easily by the respondents.
12. Avoid use of negatives (e.g., not, none, never).

Appendix K: Checklist for the Development of the Survey Items

1. Are the statements or questions in the present tense? ________
2. Did you avoid using statements that are factual or capable of being interpreted as factual? ________
3. Did you avoid statements that can have more than one interpretation? ________
4. Did you avoid statements that are likely to be endorsed by almost everyone or almost no one? ________
5. Are there almost an equal number of statements expressing positive and negative feelings? ________
6. Are the statements short? ________
7. Is each statement a proper grammatical sentence? ________
8. Did you avoid statements containing universals such as *all*, *always*, *none*, and *never*? ________
9. Did you avoid using indefinite qualifiers such as *only*, *just*, *merely*, *many*, *few*, or *seldom*? ________
10. Did you use simple sentences? ________
11. Did you avoid statements that contain "if" or "because" ________
    clauses?
12. Did you use vocabulary that can be easily understood? ________
13. Did you avoid use of negatives (e.g., *not*, *none*, *never*)? ________

Appendix L: Wrong Responses for each Frame of each Tutorial Task

Table 1

The Task Names and Task Numbers Used in Table 2

Task	Task Number
Primer	1
Primer Practice Task	2
Pretest	3
Pre-Survey	4
Graphing Proficiency Test	5
Basic Graphing	6
Basic Graphing Practice Task	7
The Control And Measurement Of Behavior	8
The Importance Of Graphing	9
Behavioral Graphing Concepts	10
The Cumulative Graph	11
The Cumulative Graph Practice Task	12
The Multiple Baseline Graph	13
The Multiple Baseline Graph Practice Task	14
Post-Survey	15
Posttest	16

Appendix L: (Continued)

Table 2

The Number of Wrong Responses for each Frame of each Tutorial Task

	Task (see Table 1 for task names)															
Frame	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
0	2	245	1	0	1	3	50	2	0	3	3	20	3	23	0	3
1	0	35	42	0	10	3	20	5	43	34	3	17	48	6	0	26
2	29	19	50	0	33	19	20	10	67	65	19	16	30	7	0	40
3	2	13	40	0	11	72	6	18	43	35	63	13	3	9	0	35
4	0	9	16		38	35	17	52	15	21	13	14	0	30	0	11
5	24		41		30	12	13	23	12	12	33	8	22	5	0	37
6	16		41		19	7	14	65	58	89	4	27	8	2	0	41
7	2		37		35	7	24	52	15	47	18	3	18	7	0	32
8	62		29		15	7	13	26	14	16	11	4	26	10	0	29
9	32		27		27	6	10	11	2	21	19	11	20	4	0	36
10	0		7		27	10	5	48	24	13	5	142	48	7	0	9
11	39		5		0	16	50	64	24	70	87	80	11	53		5
12	54		15			0	1	58	11	15	33	31	25	112		12
13	12		29			47		35	54	22	84	18	9	20		34
14	4		30			48		20	55	57	20	40	23	12		35
15	45		11			18		22	14	15	40		32	10		10

Appendix L: (Continued)

	Task (see Table 1 for task names)															
Frame	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
16	88		18			53		23	31	31	63		86	87		10
17	1		29			21		18	30	61	28		32	30		27
18	28		26			57		1	0	60	62		17	46		35
19	61		26			1		8	9	11	24		62	18		31
20	47		13					38	23	10	30		10	9		8
21	52		25					16	40	7	36		12	4		6
22	24		13					21	25	27	17		13	1		8
23	33		4					68	17	3	7		27	7		4
24	34		29					85	23	9	9		31	14		37
25	25		4					3	79	2	36		14	69		3
26			11						19	72	28		37	37		6
27			14						24	37	31		49	22		13
28			5						22	6	8		3	11		9
29			30						68	54	8		5	9		26
30			14						13	28	7			2		10
31			27						69	11	37			5		19
32			7						60	25	12			7		6
33			5						4	3	17			29		9

Appendix L: (Continued)

	Task (see Table 1 for task names)															
Frame	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
34			5						8	33	35			10		6
35			14						13	53	21			5		9
36			30						80	24	15			12		36
37			19						12	29	26			9		7
38			21						64	19	22			14		17
39			25						9	59	22			18		34
40			2						21	4	3			16		8
41									23	33				9		
42									1	19				17		
43										19				7		
44										13						
45										9						
46										29						
47										17						
48										16						
49										62						
50										50						

Appendix L: (Continued)

	Task (see Table 1 for task names)															
Frame	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
51										11						
52										44						
53										5						
54										15						
55										16						
56										20						
57										53						
58										41						
59										42						
60										7						
61										4						
62										15						
63										32						
64										4						
65										55						
66										44						
67										39						

Appendix L: (Continued)

	Task (see Table 1 for task names)															
Frame	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
68										29						
69										16						
70										25						
71										55						
72										86						
73										0						

Appendix M: The Final Version of the Items in the Knowledge Assessment

1. A Cumulative Graph can go up or down, depending on the data being used. (Type true or false)
2. The ______ ______ is a concise statement that, in combination with the axis and phase/condition labels, provides the reader with sufficient information to identify the independent and dependent variables.
3. Applied behavior analysis is characterized by the search for and demonstration of experimental ______ over socially important behavior.
4. On a Cumulative Graph, response rates are compared with one another by comparing the slope of each data path. (Type true or false)
5. The Y axis (ordinate) represents the dimension of behavior. It represents the ______ variable.
6. ______ lines in a graph usually indicate a change in treatments.
7. Experimental control is achieved when predictable and quantifiable ______ in an individual's behavior can be reliably and repeatedly produced by the systematic manipulation of some aspect of the person's environment.
8. Data points can be placed on either axes without causing distortion. (Type true or false)
9. Control is shown by a change in the independent variable when we manipulate environmental variables. (Type true or false)
10. Visual analysis focuses upon both variability in data and trends. (Type true or false)
11. Multiple data paths can be used to represent measures of the same behavior taken under different experimental conditions. (Type true or false)

12. A ______ reveals, in a more durable way, the effects of a procedure, suggests whether to try something new, or suggests whether to reinstitute a previous condition.
13. Quantitative measurement, for example surveys, is critically important in a science of behavior. . (Type true or false)
14. In order to objectively document and quantify behavior change, direct and repeated ______ of behavior is conducted.
15. In applied behavior analysis, the Pie Chart is the most common form of data display. (Type true or false)
16. ______ is the frequency of responses emitted per unit of time, usually reported as responses per minute in applied behavior analysis.
17. A data ______ is created by connecting successive data points within a given phase or condition with a straight line.
18. Gradual changes in ______ from one rate to another can be hard to detect on cumulative graphs.
19. A phase change line indicates major changes in the dependent variable. (Type true or false)
20. Making valid and reliable decisions from the raw data is extremely difficult. (Type true or false)
21. In a ______ record, the Y-axis value of any data point represents the total number of responses recorded since the beginning of data collection.
22. Behavior analysis takes the deterministic point of view. This philosophical point of view assumes that there are ______ for all events.

23. A behavior analyst must maintain direct and continuous contact with the behavior under investigation otherwise s/he cannot scrutinize it and its causes carefully. (Type true or false)

24. In a multi-tiered graph, the data from multiple individuals are often stacked ______ within the same graph.

25. In a science of behavior we manipulate environmental variables while we measure changes in behavior. (Type true or false)

26. On a Cumulative Graph, a steep slope indicates a low rate of responding. (Type true or false)

27. A ______ variable is what you measure to learn whether there is an effect in a possible cause/effect relationship.

28. On multiple-tier graphs more than one label identifying the behavioral measure is appropriate. (Type true or false)

29. ______ analysis is used to determine whether a change in behavior is consistent and related to the treatment.

30. Data points that fall on either side of a phase change line should be connected. (Type true or false)

31. A ______ may be described as a major change in reinforcement contingencies, while a "condition" may be a minor variation of that phase's contingency.

32. The X axis is a straight, horizontal line that represents the passage of ______ during repeated measures.

Appendix M: (Continued)

33. Inaccurate placement of data points is an unnecessary source of distortion in graphic displays. (Type true or false)

34. People who are unfamiliar with Cumulative Records find them hard to read because they do not go downward when behavior ceases.

35. When multiple data paths are displayed on the same graph, only one line style should be used for the data paths. . (Type true or false)

36. The ______ should use the same terms or phrases found in the textual discussion of the procedure accompanying the graph.

37. One look at the most recent data point on a ______ graph reveals the total amount of behavior up to that point in time.

38. The graphic display of ______ allows and encourages independent judgments and interpretations of the meaning and significance of behavior change.

39. Taking averages of performances during various conditions and plotting them would reveal trends in the data. (Type true or false)

40. Time is a variable in all experiments and should not be distorted arbitrarily in a graphic display. (Type true or false)

Appendix N: Itemized Summary of the Posttest Data

Item	Item Type	Correct Answers	Wrong Answers	Item Difficulty Index	r_{pb}	*p* (2-tailed)
1	Alternate-Choice	21	14	.60	.59	.00
2	Short-Answer	10	25	.29	.36	.03
3	Short-Answer	16	19	.46	.53	.00
4	Alternate-Choice	34	1	.97	-.01	.96
5	Short-Answer	7	28	.20	-.08	.67
6	Short-Answer	7	28	.20	-.04	.81
7	Short-Answer	11	24	.31	.13	.46
8	Alternate-Choice	21	14	.60	.60	.00
9	Alternate-Choice	8	27	.23	.41	.01
10	Alternate-Choice	33	2	.94	-.04	.82
11	Alternate-Choice	33	2	.94	.41	.01
12	Short-Answer	27	8	.77	.36	.03
13	Alternate-Choice	4	31	.11	.37	.03
14	Short-Answer	10	25	.29	.26	.13
15	Alternate-Choice	28	7	.80	.48	.00
16	Short-Answer	29	6	.83	.31	.07
17	Short-Answer	11	24	.31	.43	.01

Appendix N: (Continued)

Item	Item Type	Correct Answers	Wrong Answers	Item Difficulty Index	r_{pb}	*p* (2-tailed)
18	Short-Answer	5	30	.14	.36	.04
19	Alternate-Choice	8	27	.23	.29	.09
20	Alternate-Choice	31	4	.89	.27	.11
21	Short-Answer	32	3	.91	.38	.02
22	Short-Answer	32	3	.91	.50	.00
23	Alternate-Choice	34	1	.97	.19	.28
24	Short-Answer	6	29	.17	.40	.02
25	Alternate-Choice	35	0	1.00		
26	Alternate-Choice	32	3	.91	.55	.00
27	Short-Answer	26	9	.74	.02	.92
28	Alternate-Choice	30	5	.86	.09	.59
29	Short-Answer	15	20	.43	.09	.60
30	Alternate-Choice	28	7	.80	.52	.00
31	Short-Answer	22	13	.63	.47	.00
32	Short-Answer	32	3	.91	.55	.00
33	Alternate-Choice	31	4	.89	.07	.71
34	Alternate-Choice	32	3	.91	-.01	.94

Appendix N: (Continued)

Item	Item Type	Correct Answers	Wrong Answers	Item Difficulty Index	r_{pb}	p (2-tailed)
35	Alternate-Choice	29	6	.83	.45	.01
36	Short-Answer	4	31	.11	.18	.30
37	Short-Answer	33	2	.94	.19	.28
38	Short-Answer	21	14	.60	.40	.02
39	Alternate-Choice	7	28	.20	.30	.08
40	Alternate-Choice	33	2	.94	-.01	.95

Appendix O: The Skills Assessment Items

Item 1:

A teacher was concerned about a particular child's use of profanity towards other students. The teacher decided to gather data on the behavior, noting the occurrences of profanity each day for two weeks. The result was:

Day	Occurrences
1	2
2	1
3	3
4	0
5	2
6	1
7	0
8	3
9	1
10	2

Construct a Cumulative graph of the data.

Item 2:

A high school librarian was concerned about incidences of loud noises in the library during a specific time of the day. The librarian decided to try a simple solution and play classical music during this particular period.

Appendix O: (Continued)

As a baseline, the librarian collected data for 5 days, counting the number of noise incidences each day. The following week, classical music was played for 5 days and the library again counted the number of noise incidences. During the third week, classical music was not played for the first 3 days, but it was played during the last two days. The following data were collected:

(Baseline: 6, 9, 7, 9, 10)

(Classical Music: 5, 4, 4, 2, 3)

(Baseline: 8, 10, 7)

(Classical Music: 4, 3)

Construct a graph of the data.

Appendix P: The Graphing Proficiency Items

1. The horizontal axis is the ______ axis
2. The X and Y ______ of a point describe the point's location on a 2-dimensional plane.
3. ______ is the X coordinate of the origin. (Use numbers)
4. The axes of a graph are ______, this means that they form a ______ (use numbers) angle at the point where they intersect.
5. It is important to label all tick marks. (Type true or false)
6. (2,4) (is/is not) ______ a valid representation of a pair of coordinates. 5-5 (is/is not) ______.
7. The ______ coordinate system describes a point using 2 numbers.
8. A line is described by at least ______ points. (Use numbers)
9. A line segment is a part of a line bounded by 2 points. A ______ is a line that starts at 1 point and extends to infinity.
10. Tick marks can be placed at different intervals along an axis. (Type true or false)

Appendix Q: Actual Responses from in the One-To-One Evaluation Group

Question	Participant 1	Participant 2	Participant 3
1	Yes	No	No
2	Yes	No	No
3	Yes	Yes	Some what
4	(no reaction)	Very effective	Very effective
5	The pace was fine	Steady pace	Slow
6	Yes	Yes	Yes
7	Yes	Yes	Yes
8	My overall reaction was misunderstood… (remaining reaction refers to the course content)	At start a little difficult to understand	Good
9	None really	Helps user by allowing user to reinput correct reply	Detailed
10	Not enough details on why graphing was important	Directions could be clearer/easier to understand	Time consuming

Appendix R: A Summary of the Verbal Responses from Participants in the One-To-One Evaluation Group

Topic	Participant 1	Participant 2	Participant 3
Interface	1. Text too small. 2. Difficult to graph with mouse. 3. Some options were shaded and confusing.	Unclear, not sure what to do next, but it became clear after a while.	1. Usable but crowded. 2. No time for freelance.
Content	1. Good but not enough details. 2. Learn some things from the tutorials 3. Too long.	Content was not hard.	1. Hard 2. Spelling errors 3. Pictures need to remain visible (no scrolling)
Functionality	Freelance was not mandatory so it was avoided	Program did not work initially.	1. Overview needed 2. Make instruction precise and concise.
Likes	1. Informative. 2. Opportunity to actually graph. 3. Graphing screen showed when the answer was wrong.	Review Screen after incorrect answers.	Tutorials move from one to the next easily.

Appendix R: (Continued)

Topic	Participant 1	Participant 2	Participant 3
Dislikes	1. Instruction text size was too small. 2. Seeing the frame score was intimidating. 3. Took too long. 4. Would prefer more modules and fewer frames per module.	Too long, would prefer more tutorials and fewer frames.	Lay words in the tutorials.
Changes	1. Increase text sizes. 2. No scrolling.	1. Increase text size. 2. clarify options.	1. Add an overview. 2. Streamline instruction.

Appendix S: A Summary of the Small-Group Evaluation Responses

1. Did you have any problems with specific words or phrases?

Ten participants reported no problems with specific words or phrases. The 3 participants who reported having problems did not specify the words or phrases that caused the problems.

2. Did you have any problems with specific sentences?

Ten participants reported no problems with specific sentences. The three participants who reported having problems did not specify the sentences that caused the problems. One participant noted that a few sentences had errors, and another learner noted that some sentences could have been more precise.

3. Did you understand the themes presented in the instruction?

Twelve participants reported that they understood the themes, while one participant reported that they only somewhat understood the themes.

4. How effective was the sequencing of the instruction?

All 13 participants reported that the instruction was well sequenced and effective. One participant mentioned that the instruction was too long and repetitive, and another participant noted that it was not bad but could have been better. The participant did not specify in what ways the instruction could have been better.

5. How would you describe the delivery pace of the instruction?

Eleven responses described the instruction pace as good. The 2 participants who did not report the pace as being good noted that the instruction was at times choppy and a little vague, and that there were too many frames. Four participants who reported that the pace was good also mentioned that the instruction was too long.

6. Was sufficient content presented in each section?

Eleven participants reported that there was sufficient content in each section, while the remaining two participants were uncertain about either the question or their responses to the question.

7. Was there sufficient time to complete the instruction?

Twelve participants reported that they had enough time to complete the instruction, while one participant reported that the time was not sufficient.

8. What is your overall reaction to the instruction?

All 13 participants reacted positively to the instruction, reporting that the instruction was effective and organized. In addition to the reacting positively, 6 participants added that the instruction was too lengthy.

9. What are specific strengths of the instruction?

Specific strengths included: vocabulary and detailed explanations, clarity and emphasis on terms, allows active working, step-by-step instruction, simple and easy, the ability to repeat modules, and the use of repetition and pictures.

10. What are specific weaknesses of the instruction?

Specific weakness included: length, some vague sentences, animation and color, and some unclear instruction. Most participants reported that the instruction was too long. One participant noted that they would have preferred fewer frames and more modules.

Appendix T: LOGIS Java Code

Email the author for the LOGIS Java code.

About the Author

Darrel Davis was born in Belize City, Belize but he spent much of his childhood in the capital city Belmopan. He completed his Bachelor of Science in Mathematics Education at the University College of Belize then earned a Master of Science in Computer Science at the University of South Florida. His interest in technology and teaching led him to the Instructional Technology program in the College of Education at the University of South Florida

Darrel has taught at the high school level and at the university level. He has taught undergraduate students and graduate students, and has extensive experience with online learning.

Mr. Davis is currently a Heanon Wilkins Fellow at Miami University of Ohio.