Formative Research on an Instructional Design Model for the Design of Computer Simulation for Teaching Statistical Concepts

A dissertation presented to
the faculty of
the College of Education of Ohio University

In partial fulfillment
of the requirements for the degree
Doctor of Philosophy

Chung-Yuan Hsu
November 2009

© 2009 Chung-Yuan Hsu. All Rights Reserved.
This dissertation titled
Formative Research on an Instructional Design Model for the Design of Computer
Simulation for Teaching Statistical Concepts

by

CHUNG-YUAN HSU

has been approved for
the Department of Educational Studies
and the College of Education by

David R. Moore
Associate Professor of Educational Studies

Renée A. Middleton
Dean, College of Education
ABSTRACT

HSU, CHUNG-YUAN, Ph.D., November 2009, Curriculum and Instruction, Instructional Technology

Formative Research on an Instructional Design Model for the Design of Computer Simulation for Teaching Statistical Concepts (203 pp.)

Director of Dissertation: David R. Moore

The instructional problem of a superficial understanding still prevails in current education. Many educators seek solutions from technology to remedy the shadow learning problem. But, as researchers indicate, technology alone does not cause learning. Rather, learning is influenced more by instructional interventions.

An instructional design model that fulfills the aim of meaningful learning is Goal-Based Scenarios (GBSs) proposed by Schank, Fano, Bell, and Jona (1993/1994). It offers guidelines to guide the design of a computer simulation. Despite growing evidence that supports the effectiveness of using the GBS model, no empirical studies have investigated strengths, weaknesses, or possible improvement of the GBS model. Thus, the purpose of the present study is to evaluate the GBS model by answering following questions: 1) what are the strengths and weaknesses of the GBS model? 2) What improvements can be made?

Formative research was employed to investigate the designed instance by using think aloud interview, debrief (semi-structured) interview, and a focus group interview. The result showed that a GBS might become a better instructional design model if improvements are made in these aspects: 1) provide a worked example or instruction that demonstrates the behaviors of using GBS and seeking supports in order to increase the
user’s lower sense of self-efficacy while pursuing mission or assuming the role, 2) employ approaches of a small group usage and open-ended question to promote learners’ engagement and interaction in scenario operations, 3) carefully integrate other components in GBS to support hands-on activity, 4) provide cues in negative feedback and recapitulate the concept in positive feedback.

Approved: ________________________________________________________________

David R. Moore

Associate Professor of Educational Studies
ACKNOWLEDGMENTS

I cannot express enough thanks to the members of my committee. This dissertation would not have been possible to complete without their guidance and encouragement. I am very appreciative of all the time and effort they had offered me.

Special thanks to:

Dr. David Moore (my helpful and approachable advisor)

Dr. George Johanson

Dr. Teresa Franklin

Dr. Francis Godwyll

I appreciate the assistance and inspiration from my colleagues, Gerard Akindes, Ken Dobo, Mark Riley, Douglas Franklin, and Matt Ziff. I am thankful to Brendan Morse, Dr. Johanson, and Dr. Gordon Bruce who helped me develop the instructional system.

Thanks also to my friends, Jerry Yang, Yvonne Chou, Marje Lang, David Dominguese, and Sophia Pan for friendship and support.

Finally, I want to dedicate this dissertation to my wife Kai-Lin Lin and my parents who supported and encouraged me through many years of study. There are so many friends in Athens and Taiwan who helped me. If you are not mentioned here, please know that I am grateful for what you have done for me.
# TABLE OF CONTENTS

<table>
<thead>
<tr>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Abstract</td>
</tr>
<tr>
<td>Acknowledgments</td>
</tr>
<tr>
<td>List of Tables</td>
</tr>
<tr>
<td>Chapter 1: Introduction</td>
</tr>
<tr>
<td>Background of the Study</td>
</tr>
<tr>
<td>Statement of the Problem</td>
</tr>
<tr>
<td>Significance of the Study</td>
</tr>
<tr>
<td>Delimitations and Limitations of the Study</td>
</tr>
<tr>
<td>Definition of Terms</td>
</tr>
<tr>
<td>Organization of the Study</td>
</tr>
<tr>
<td>Chapter 2: Literature Review</td>
</tr>
<tr>
<td>Introduction</td>
</tr>
<tr>
<td>Theoretical Foundation</td>
</tr>
<tr>
<td>Constructivism</td>
</tr>
<tr>
<td>Situated Cognition</td>
</tr>
<tr>
<td>Motivation and Learning</td>
</tr>
<tr>
<td>Pretask</td>
</tr>
<tr>
<td>During Task</td>
</tr>
<tr>
<td>Posttask</td>
</tr>
<tr>
<td>Computer Simulations</td>
</tr>
<tr>
<td>Computer Simulations for Statistics Education</td>
</tr>
</tbody>
</table>
Design an Instance of the Theory ................................................................. 62
Collect and Analyze Formative Data on the Instance................................. 67
Revise the Instance ...................................................................................... 73
Repeat the Data Collection and Revision Cycle ......................................... 74
Offer Tentative Revisions for the Theory ..................................................... 74
Methodological Issues .............................................................................. 74
   Construct Validity .................................................................................. 74
   Sound Data Collection and Analysis Procedures ...................................... 76
   Attention to Generalizability to the Theory .............................................. 78
Participants .............................................................................................. 78
   The Pilot .................................................................................................. 80
   Summary ................................................................................................. 81
Chapter 4: Findings .................................................................................. 82
   Introduction ........................................................................................... 82
   Learning Goal ......................................................................................... 82
      What Happened? ................................................................................ 82
      What Went Well and What Didn’t? .................................................... 83
   Mission .................................................................................................. 85
      What Happened? ................................................................................ 85
      What Went Well and What Didn’t? .................................................... 85
   Cover Story ............................................................................................ 87
      What Happened? ................................................................................ 87
      What Went Well and What Didn’t? .................................................... 88
<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Role</td>
<td>90</td>
</tr>
<tr>
<td>What Happened?</td>
<td>90</td>
</tr>
<tr>
<td>What Went Well and What Didn’t?</td>
<td>90</td>
</tr>
<tr>
<td>Scenario operations</td>
<td>93</td>
</tr>
<tr>
<td>What Happened?</td>
<td>93</td>
</tr>
<tr>
<td>What Went Well and What Didn’t?</td>
<td>94</td>
</tr>
<tr>
<td>Resources</td>
<td>96</td>
</tr>
<tr>
<td>What Happened?</td>
<td>96</td>
</tr>
<tr>
<td>What Went Well and What Didn’t?</td>
<td>96</td>
</tr>
<tr>
<td>Feedback</td>
<td>99</td>
</tr>
<tr>
<td>What Happened?</td>
<td>99</td>
</tr>
<tr>
<td>What Went Well and What Didn’t?</td>
<td>101</td>
</tr>
<tr>
<td>Results of Understanding</td>
<td>102</td>
</tr>
<tr>
<td>Summary</td>
<td>112</td>
</tr>
<tr>
<td>Chapter 5: Discussion and Conclusion</td>
<td>113</td>
</tr>
<tr>
<td>Strengths of GBS</td>
<td>114</td>
</tr>
<tr>
<td>Learning Goals</td>
<td>114</td>
</tr>
<tr>
<td>Mission</td>
<td>114</td>
</tr>
<tr>
<td>Cover Story</td>
<td>115</td>
</tr>
<tr>
<td>Role</td>
<td>115</td>
</tr>
<tr>
<td>Scenario Operations</td>
<td>116</td>
</tr>
<tr>
<td>Resources</td>
<td>117</td>
</tr>
<tr>
<td>Feedback</td>
<td>118</td>
</tr>
</tbody>
</table>
LIST OF TABLES

Table 1: Framework of Statistics Specialist ................................................................. 63
Table 2: Demographic Information of Participants ..................................................... 79
Table 3: Participant Distribution of Mission and Role Awareness .............................. 86
Table 4: Participants’ Suggestions for Resources ...................................................... 100
Table 5: Time Spent on Resources .......................................................................... 105
Table 6: Descriptive Information for Time Spent on Resources .............................. 105
Table 7: Results of Pretest ...................................................................................... 106
Table 8: Results of Posttest .................................................................................... 106
Table 9: Participants’ Explanations to Question 1 in the Retention Test .................. 108
Table 10: Participants’ Explanations to Question 2 in the Retention Test ................. 109
Table 11: Participants’ Explanations to Question 3 in the Retention Test ................. 110
Table 12: Participants’ Explanations to Question 4 in the Retention Test ............... 111
List of Figures

<table>
<thead>
<tr>
<th>Figure</th>
<th>Description</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Figure 1</td>
<td>The learning path of Goal-Based Scenarios Computer Simulation</td>
<td>66</td>
</tr>
<tr>
<td>Figure 2</td>
<td>Introduction of Role-Playing</td>
<td>121</td>
</tr>
</tbody>
</table>
CHAPTER 1: INTRODUCTION

Background of the Study

Over the past few decades there has been a strong movement towards education that moves beyond the simple rote learning common to the industrial age, toward meaningful learning suited for the information age (Reigeluth, 1999).

According to Ausubel (2000), meaningful learning refers to learned knowledge that is fully understood by the individual and that the individual knows how that specific fact relates to other stored facts. Consistent with meaningful learning is the constructivists’ paradigm wherein learning is knowledge construction in which individuals seek to make sense of their experiences from their own unique cognitive perspective (Ausubel, 2000; Duffy & Jonassen, 1992; Gredler, 2009; Mayer, 1999; von Glasersfeld, 1987).

Kintsch (2009) stated that meaningful learning occurs when an individual turns the input (e.g. written text or other instructional media) into the end of the process as evidenced by a situation model that represents the meaning of the media presented. It is an active and constructive process, regardless of how the input cues students during the learning activity. It is the students’ individual construction of the situation model in their mind that makes the learning activity meaningful (Kintsch, 2009). Since students may hold deeply rooted misconceptions about the targeted knowledge, researchers (Duit & Treagust, 2003; Jonassen, Strobel, & Gottdenker, 2005) of conceptual change state that meaningful learning occurs during the process of constructing and reorganizing individual conceptual models through experience and reflection.
The instructional problem of a superficial understanding still prevails in current education. Take statistics education for example. Many teachers and researchers perceive that students who pass a statistics class have a shallow and isolated understanding of foundational concepts and have difficulty applying these concepts to reasoning (Chance, delMas, & Garfield, 2004; delMas, Garfield, & Chance, 1999). These students may be able to do the necessary calculations but unable to comprehend the underlying process. Without deep comprehension, students in later classes tended to use a rote manipulation approach for statistical inference and were unable to interpret research studies accurately (Chance et al, 2004). This superficial or isolated understanding of foundational concepts is also known as shadow learning.

Many educators seek solutions from technology to remedy the shadow learning problem. Bransford (2000) indicated that the use of simulations may “engage learners as active participants in their learning by focusing their attention on critical elements, encouraging abstraction of common themes or procedures (principles), and evaluating their own progress toward understanding” (p. 68). Nickerson (1995) revealed that while learning in simulations, individuals can perform tasks in a supportive environment to construct their own meanings. Simulations not only quicken the random and complex learning processes that may take a long time to display in the real world, but also provide opportunities to conceptualize and test ideas by experimenting parameters (Mills, 2004; Windschitl & Andre, 1998; Yu & Behrens, 1995). Achieving these merits requires an essential condition: the simulations must be carefully designed.

As Shon (1996) pointed out, “Technology alone does not cause learning. Rather, learning is influenced more by instructional interventions” (p. 3). Reigeluth (1999) stated
that with an instructional design model, learning is enhanced by identifying specific events based on learning theories to facilitate and support learning, recognizing situations where specific methods should or should not be used, and increasing the chances of achieving the instructional goals.

An instructional design model that fulfills the aim of meaningful learning is Goal-Based Scenarios (GBSs) proposed by Roger Schank, Andrew Fano, Benjamin Bell, and Menachem Jona (1993/1994). The model was later revised by Schank, Berman, and Macpherson (1999). This GBS model offers guidelines to guide the design of a computer simulation. Within a simulated environment, students are responsible for achieving a goal by assuming a role in the scenarios and seeking different learning paths to collect information and complete the mission (Schank et al., 1993/1994). In addition, video clips are often used as a feedback method to communicate the consequence of their actions. The GBS model consists of seven essential components (Schank et al., 1999):

1. Learning goal: The goal is the teaching points that the designer or educator would like students to acquire or improve.

2. Mission: The mission refers to the task that requires knowledge or skills that the designer would like to impart, in order to achieve the goal.

3. Cover story: The cover story is the storyline creating opportunities for the mission to be accomplished.

4. Role: The role refers to the character that the students assume while pursuing the mission.

5. Scenario operations: Scenario operations include activities that the learners perform in pursuit of the mission.
6. Resources: Resources are essential information provided to help students complete the task.

7. Feedback: The feedback is given through the consequences of actions.

The purpose of instructional design is to facilitate the process of learning rather than the process of teaching (Gagné, Wager, Golas, & Keller, 2005). It aims to foster an intentional learning that is guided by the target goal and the desired learning outcome (Gagné et al., 2005). Researchers in the field of instructional design have concerned themselves with the design of computer simulations and have proposed models to guide the design task. However, more empirical studies are needed to validate instructional design theories or models. As Gagné et al. indicated, design is an iterative process. Instructional models and materials must be evaluated by learners and designers to verify what works and what does not. This process is described by Gagné et al. (2005), “Designers don’t design perfect instruction; they perfect instructional designs” (p. 3).

Statement of the Problem

Despite more evidence that supports the effectiveness of using the GBS model (Bell, Bareiss, & Beckwith, 1993/1994; Campbell & Monson, 1994; Collins, 1994; Iverson & Colky, 2007; Naidu, Ip, & Linser, 2000. Schank et al., 1993/1994; Schank et al., 1999; Schoenfeld-Tacher, Jones, & Persichitte, 2001), no empirical studies have investigated strengths, weaknesses, or possible improvement of the GBS model. Thus, the purpose of the present study is to evaluate the GBS model by answering following questions:

1. What are the strengths and weaknesses of the GBS model?

2. What improvements can be made?
Formative research (Reigeluth & Frick, 1999), a qualitative methodology designed to improve instructional design models, was used in the present study to assess the GBS model. The two study questions were answered by examining participants’ responses during and after an instructional instance that was designed in terms of the GBS model. Since an effective instructional design model facilitates the development of an instructional tutorial that promotes students’ understanding, the participants’ understanding of the targeted knowledge (Central Limit Theorem) was assessed as well. This assessment took the form of an evaluation of participant responses in the posttest and retention test.

Significance of the Study

This study is significant in that it is the first empirical evaluation of the GBS instructional design model. Results of this research will extend the knowledge base for the GBSs model and provide insights into the development and application of computer simulations in statistics education. Further, the results of this study can offer useful information to inform researchers, educators, or instructional designers who are interested in using computer simulation to facilitate statistical learning and teaching. Finally, with no published studies focusing on an empirical approach to evaluate a GBS model, the results of the present study contribute to filling a gap in the current body of research.

Delimitations and Limitations of the Study

Understanding the boundaries of the investigation is very important. The purpose of this study was to evaluate an instructional model, Goal-Based Scenarios. Participants were restricted to midwestern university students. The limitations of this study included:
1. This study was limited to investigate the instance, *Statistics Specialist*, that was developed by the researcher following the GBSs model to teach sampling distributions.

2. Participants were limited to graduate level students who were restricted to one usage of *Statistics Specialist*, which took approximately one hour to complete.

3. This study was limited to information acquired from interviews and field observations. Interviewee responses and researcher observations are subject to unknown preconceptions.

4. The findings only represent the experiences of the students who used *Statistics Specialist* to learn statistical concepts and cannot be applied to the content of different domains. Moreover, since the design of Statistics Specialist was based on the GBSs model, the findings also cannot be generalized to different instructional models.

5. The present study represents a single iteration of formative research. Possible strengths, weaknesses, and recommendations were identified from data collection. A GBS model could be better evaluated with additional cycles of formative research, which may correct identified problems and influence emerging data to better understand implications for the model (Watson, 2007).

**Definition of Terms**

*Computer simulations:* Computer simulations, also called digital simulation, refer to a computer-implemented program designed to teach someone about a modeling system or process that can be either natural or artificial (Baek, 2009). Some essential features of simulation include (Gredler, 2004): 1) sufficient opportunities for a student to interact
with the context; 2) a defined role with responsibilities and constraints; 3) a data-rich environment allowing users to perform a range of strategies; and 4) feedback to support problem-solving and act as results of consequences. For educational purposes, high fidelity is not necessary to make for a better simulation. Simulation must have a correct level of fidelity since overly simplified representations may confuse learners and exact representations may make them too complex (Blake & Scanlon, 2007).

*Formative research:* Formative research proposed by Reigeluth and Frick (1999) is “a kind of developmental research or action research that is intended to improve design theory for designing instructional practices or processes” (p. 633). Due to the fact that quantitative research methodologies are not helpful to the early stage of instructional design theoretical development, researchers tend to use qualitative research methods to evaluate how the theory meets three essential criteria: effectiveness, efficiency, and appeal (Reigeluth & Frick, 1999).

*Sampling distributions:* Sampling distributions are the most fundamental concept underlying all statistical tests (Howell, 2002). If a large sample is obtained, three rules of sample distribution will follow: 1) the mean of a distribution of means is the same as the mean of the population; 2) the variance of a distribution of means is the variance of the population divided by the number of sample size; 3) the shape of a distribution of means is approximately normal if either the size of each sample is more than thirty or the distribution of the population is normal (Aron, Aron, & Coups, 2005). These three rules are also called the Central Limit Theorem (CLT).
Organization of the Study

This dissertation consists of five chapters. Chapter one introduces the background of the study, the problems, and research questions. Chapter two provides a theoretical review of the literature supporting this study. Chapter three describes the methods and procedures used in this study, including the development of an instructional instance, Statistics Specialist, and the results of pilot studies. Chapter four consists of a presentation of the results. Chapter five discusses the implications of the findings and offers possible recommendations for future research.
CHAPTER 2: LITERATURE REVIEW

Introduction

This chapter reviewed the theoretical foundation and research literature supporting this study. Discussed topics include: learning theories underlying a GBS model, motivation and learning, computer simulation and its application in statistics education, characteristics of GBS and current studies on GBS, distinctions among problem-based learning, case-based learning, and GBS, distinctions between simulations and games design model, formative evaluation, and the developmental model on sampling distributions.

Theoretical Foundation

Learning theories describe how learning takes place. An investigation of learning theories not only helps the researcher understand how an instructional design model works but also assists the researcher in evaluating and improving more thorough aspects of the instructional design model (Reigeluth, 1999). Learning theories such as constructivism and situated cognition share similar characteristics. Together, they provide perspectives on learning that greatly influence the GBS model.

Constructivism

Constructivism, the philosophy of understanding (Goodman, 1984) and a theory of knowing (von Glasersfeld, 1995), claims that “…. reality [knowledge] is more in the mind of the knower, that the knower constructs a reality, or at least interprets it, based upon his or her apperceptions” (Jonassen, 1991b, p. 10). Constructivism is often described as student-centered, learner-centered, self-directed, and exploratory learning. Regardless of different labels, all of these terms share the same critical factor, that a
learner’s mental activity plays an important role in their knowledge construction (Cunningham, 1992; Jonassen, 1991b). Generally, constructivists do agree on five basic characteristics of constructivism.

First, knowledge is not transmitted but individually constructed by learners (Bednar, Cunningham, Duffy, & Perry, 1992; Ertmer & Newby, 1993; Jonassen, 1999). No matter how clearly an instructor explains a concept, learners comprehend the concept only when they have constructed their own meaning (Jonassen, 1999). This is an internal process that requires students to independently reinterpret and reorganize the new information so they can “buy in” and link it to their own existing knowledge or experiences (Perkins, 1993).

Second, knowledge is individualistic (Jonassen, 1991a; von Glasersfeld, 1995) and its construction is not equally viable (Savery & Duffy, 1995). Since learners can only interpret information in the context of their own experiences and interactions, what they comprehend and interpret will be, to some extent, individualistic, only accessible to them, and true to themselves but not anyone else.

Third, knowledge constructed is constantly evolving (Bednar et al., 1992; Ertmer & Newby, 1993). The knowledge one constructs does not stay intact; it evolves more or less with each new use in different situations, social negotiations or interactions (Driscoll, 2005; Jonassen, 1999), and activities.

Fourth, knowledge evolves through social construction (Piaget, 1973; Wadsworth, 2004). Savery and Duffy (1995) stated that social interactions play an important role in learning activities “we can test our own understanding and examine the understanding of others as a mechanism for enriching, interweaving, and expanding our understanding of
particular issues or phenomena” (p. 136). As von Glasersfeld (1989) stressed, it is the interaction with others that provides the most frequent source of alternative views to challenge developing cognitive understanding and stimulate learning.

Learning becomes effective when it takes place in relevant context (Cunningham, 1992; Jonassen, 1991b). According to Jonassen (1991b), since knowledge is constructed by experiences, the context or situations to which knowledge is related provides the authenticity of experiences that are critical to promote learning. Context is an integral part of what is learned and “… understanding is developed through continued, situated use” (Brown, Collins, & Duguid, 1989, p. 33).

The roots of constructivism came from Piaget’s structuralism that was a complex blend of biology, epistemology, philosophy, and psychology (O'Loughlin, 1992). Influenced by his early training and work as a biologist, Piaget suggested that the intellectual development in human beings is similar to environmental adaptation observed in organisms, consisting of assimilation, accommodation, and equilibration (Wadsworth, 2004).

Assimilation, “the filtering or modification of the input” (Piaget & Inhelder, 1969, p.6), is the cognitive process that integrates external elements into existing experiences, which account for a quantitative change. Accommodation, “the modification of internal schemes to fit reality” (Piaget & Inhelder, 1969, p.6), is the process of adjusting old experiences or the creation of new experiences, which accounts for a qualitative change in cognition structure.

Piaget proposed that there should be a cognitive balance between assimilation and accommodation because too much assimilation may cause an inability to detect
differences in objects, and too much accommodation might make it difficult to detect similarities (Wadsworth, 2004). The state of balance, according to Piaget and Inhelder (1969), is called equilibrium. The state of imbalance is named disequilibrium (Wadsworth, 2004). Equilibration is “a series of active compensations on the part of the subject in response to external disturbances and an adjustment” (Piaget & Inhelder, 1969, p.159), that move from disequilibrium to equilibrium by self-regulation. This is where learning and acquisition of experiences take place.

Piaget (1973) suggested that the role a teacher plays during students’ knowledge construction is to encourage, stimulate, and support exploration. He wrote:

> It is obvious that the teacher as organizer remains indispensable in order to create the situations and construct the initial devices which present useful problems to the child. Secondly, he is needed to provide counter-examples that compel reflection and reconsideration of overhasty solutions. What is desired is that the teacher cease being a lecturer satisfied with transmitting ready-made solutions; his role should be that of a mentor stimulating initiative and research. (p. 16)

In agreement with Piaget’s ideas about teacher facilitation of student intellectual development, Vygotsky (1978) proposed that learners’ mental development occurs on two levels: the actual developmental level and the zone of proximal development. The former refers to the level at which learners are capable of solving problems independently, whereas the latter describes the level of, “potential development as determined through problem solving under adult guidance or in collaboration with more capable peer” (Vygotsky, 1978, p. 86). An instructor’s task is to provide resources, such as modeling of
knowledge or activities or promoting social interaction, for students to learn things they could not learn on their own. Vygotsky (1978) believed that “…what is in the zone of proximal development today will be the actual developmental level tomorrow” (p. 87).

In summary, constructivists view knowledge as an adaptation process of cognitive construction in which a learner in disequilibrium strives to return to equilibrium. Learning within a constructivist paradigm allows students to generate disequilibrium, then seek equilibrium through active methods (to assimilate and accommodate), while offering abundant teacher-mediated resources and support scaffolding. This conceptual framework underlies the GBS model. A GBS learning environment attempts to create rich experiences that encourage students to learn by doing. These experiences may include activities that allow learners to pursue a learning goal by completing a mission, engage in learning by assuming a role in a scenario, and solve problems with supports.

**Situated Cognition**

Situated cognition emphasizes the relationship between a learner and a situation (Schunk, 2008). The word, situated, describes the critical component of this learning concept from the constructivist perspective - that knowledge is situated through experience (Barab & Duffy, 2000). Researchers of situated cognition believe that learning becomes effective when it takes place in the context that the knowledge is raised and used, and that instruction should provide progressive and supported practice in cognitive skills (Brown et al., 1989; Driscoll, 2005). This idea implies two essential aspects of learning: knowing and doing.

Brown et al. (1989) stated that knowing and doing are reciprocal since knowledge is situated and increasingly constructed through activities. It is participation in practice...
that constitutes learning and understanding (Barab & Duffy, 2000; Greeno, 1998).

According to Collins (1991), students are more able to apply their knowledge while dealing with real problems and situations.

Situated cognition may be implemented in practice through a cognitive apprenticeship approach (Brown et al., 1989). The fundamental idea behind an apprenticeship is that an expert shows an apprentice how to do a task and then helps the novice to do it. An apprenticeship model helps focus the importance of learning activities in knowledge acquisition and stresses the context-dependent and situated nature of learning (Brown et al., 1989). This approach offers several advantages in that learners can see the process of work, involve their learning in physical activity, and understand the nature of expert practice (Collins, Brown, & Holum, 1991).

However, there are differences between traditional apprenticeships and cognitive apprenticeships. Collins, Brown, and Newman (1989) indicated that in cognitive apprenticeships, tasks and problems are selected to illustrate methods and allow learners to practice skills in various settings and enhance the complexity of tasks. That is, tasks change according to learning demands. Conversely, in traditional apprenticeships, job demands dictate apprentice tasks.

To translate the model of traditional apprenticeship to cognitive apprenticeship, Collins et al. (1991) proposed that instructors identify the processes and tasks that experts use and make them visible to students. In addition, teachers should situate abstract tasks in authentic contexts for students to understand the relevance of the work. Finally, it is essential to provide a diversity of situations and articulate the similarities between situations so that students are able to transfer what they have learned.
Collins et al. (1991) provided six apprenticeship-like approaches to integrate the concept of cognitive apprenticeship into instruction: modeling, coaching, scaffolding, articulation, reflection, and exploration. Modeling refers to an expert’s performing a task and externalizing his or her cognitive activities for learners to observe and construct a cognitive model of the process. Coaching consists of observing students’ practice and giving them instant feedback or hints to help them develop expertise. Scaffolding means that supports are offered to help students perform parts of the work and gradually fade out while learners acquire the skills. Articulation refers to methods that facilitate students articulating what they have learned or why and how they have done. Reflection occurs when learners compare their problem-solving processes with experts or peers in order to observe and learn from the nuance. Exploration refers to the extension of learning goals for students to find out their interest and practice what they already learned, which may promote autonomy.

In summary, constructivism is viewed as an internal and cyclical learning process in which learners actively organize and interpret new knowledge based on their experiences and understanding. Situated cognition emphasizes participation in practice within the context where knowledge is raised and used. Both theories regard learning as a constructive process with instruction that should facilitate exploration and provide sufficient resources to support students’ meaning making. The next section will introduce variables of motivation that influence learning.

Motivation and Learning

Motivation plays a critical role in learning by influencing learning behaviors and experiences (Rebolledo-Mendez, Boulay, & Luckin, 2006). According to Schunk,
Pintrich, and Meece’s (2008) definition, “Motivation is the process whereby goal-directed activity is instigated and sustained” (p. 4). Since motivation is viewed as a process, it is more appropriate to infer it from actions and verbalization (Schunk et al., 2008). To describe how motivation arises from thoughts and beliefs and influences learning, Schunk (2008) proposed a cognitive model that depicts the changing role of motivation during learning, including pretask, during task, and posttask.

**Pretask**

Variables that influence students’ incoming motivation for learning consist of goals, self-efficacy, and outcome expectations (Schunk, 2008). Self-efficacy is equivalent to confidence (Seifert, 2004) and refers to people’s belief about their capability to learn or do something (Bandura, 1993). Students who are highly efficacious tend to exert effort and persist in a task while dealing with difficulty, whereas students who perceive themselves incapable may generate self-doubt and become less persistent (Bandura, 1986; Schunk, 2008).

Closely related to self-efficacy is outcome expectations, which refer to anticipation of possible consequences of given actions (Bandura, 1986, 1993). As Bandura (1986) explained, “The belief that one can high jump six feet is an efficacy judgment; the anticipated social recognition, applause, trophies, and self-satisfactions for such a performance constitute the outcome expectations” (p. 391). However, positive outcome expectations alone do not guarantee high motivation (Schunk, 2008). Students are more engaged in learning when they expect the positive outcomes together with belief in their capability to attain the task.
Researchers of goal theory contend that goals reflect people’s purposes and motivate them to exert the effort necessary to meet the task demands and sustain over time (Schunk et al., 2008). According to Bandura (1986), motivation relies on goal-setting and self-regulation and “By making self-satisfaction conditional on a selected level of performance, individuals create their own incentives to persist in their efforts until their performances match internal standards” (Bandura, 1986, p. 467).

Goals also direct one’s attention to relevant task features, which affect how they process information (Bandura, 1986; Schunk et al., 2008). Researchers (Ames & Ames, 1984; Bandura, 1986) indicated that goals themselves do not effect motivation and learning but through their properties: specificity, difficulty, and proximity. As Ames and Ames (1984) pointed out, specific goals are more likely to enhance learning and motivation than general goals because they are explicit and easier to evaluate. Moderately challenging goals are more motivational than easier ones because they build a sense of efficacy and provide more learning information. Finally, proximal goals result in greater motivation than distant goals because learners can observe their progress toward short-term goals and tend to believe their capability in future learning.

Different types of goals may influence students’ engagement in learning. According to Dweck & Leggett (1988), individuals pursuing learning goals that focus on acquiring capabilities tend to consider effort (or other internal and controllable variables) as the cause of their success or failure. They believe that intelligence is malleable and that difficulty is a challenge (Dweck & Leggett, 1988; Seifert, 1995). They are persistent and have a greater preference for challenge (Dweck, 1986). On the other hand, students pursuing performance goals are apt to concern how their performance relates to others
and how others perceive them (Seifert, 2004). They consider that ability is the cause of
success and failure and that intelligence is a fixed entity (Dweck & Leggett, 1988). They
will take a challenge and maintain their effort only when their confidence is high
(Dweck, 1986).

During Task

Student motivation during learning is influenced by instructional, contextual, and
personal variables (Schunk, 2008). Contextual variables refer to social and environmental
resources. Factors such as classroom settings and peers can increase or impede
motivation for learning. According to Nolen’s (2003) study, classrooms that focus on
understanding and independent thinking have a positive relationship with students’ self-
reported satisfaction with learning. Ryan (2000) indicated that pressure from peer groups
can influence adolescent motivation in school; students’ perception of peer pressure
directly links to their self-reported attitudes regarding school.

Personal variables mainly refer to individual dispositions. Payne, Youngcourt, and
Beaubien (2007) investigated how personal traits (neuroticism, extraversion, openness,
agreeableness, and conscientiousness) influence goal orientation. They found that
students who were conscientious, extraverted, and open displayed the strongest learning
goal orientation, whereas those who showed high neuroticism and low extraversion
tended to experience a fear of failure and avoid performance goals.

Instructional variables consist of pedagogy, feedback, materials, and equipment
(Schunk, 2008). Typically, these variables are regarded as learning influences, but they
also affect motivation. For instance, prompting interesting and relevant questions as a
way to generate cognitive conflict or puzzlement can create stimulus for learning (Savery
& Duffy, 1995). Students are motivated to reconstruct their knowledge when they encounter and experience a conflict with their predictions (Wadsworth, 2004). Further, allowing manipulating parameters to affect the problems and providing formative feedback can help learners get involved in learning activities (Hmelo-Silver, 2004; Jonassen, 1999).

According to Meyer and Turner (2002), scaffolding can promote students’ motivation by “helping students build competence through increased understanding, engaging students in learning while supporting their socioemotional needs, and helping students build and exercise autonomy as learners”, (p 18). In addition, providing an authentic context that reflects the same type of cognitive challenges as those faced in real life may enhance students’ engagement in learning activities (Brown et al., 1989; Herrington & Oliver, 2000; Hmelo-Silver, Duncan, & Chinn, 2007; Jonassen, 1999). Finally, using a role-play simulation can promote intrinsic motivation. For example, DeNeve and Heppner (1997) compared undergraduate students’ response to role-playing simulations and traditional lecture and found that students described the role play simulation as helpful and they tended to recall more information from the role play simulation. However, Sturges, Maurer, and Cole’s (2009) research of an undergraduate science class studying protein analysis indicated that there was no significant difference in student performance between the role-play group and the lecture group. The advantage of the role play group over the lecture group was seen in measures of student involvement and satisfaction.
Posttask

Posttask refers to the time during and after the task is completed when students pause to reflect on their work (Schunk, 2008). The major factor that affects task engagement at this phase is attribution. Attribution refers to a person’s perceived explanation of an outcome (Seifert, 2004), such as luck, ability, peers, or environment. Researchers (Schunk et al., 2008; Seifert, 2004) indicated that students’ attributions can trigger emotions and further influence motivation. Weiner (1985) proposed a model that defined attribution in terms of three characteristics: locus (Is the observed outcome caused by the factor within the individual, such as effort?), stability (Are attributions stable and lasting or changing?), and controllability (Is the individual able to control the cause?). Students who attribute a successful performance to internal, stable, and controllable causes are more motivated than those who attribute their success to external, less stable and uncontrollable causes (Brophy, 2004; Schunk, 2008).

Variables of these three stages are cyclic; that is, each factor influences future motivation and learning. Bandura (1986) indicated that students’ motivation tends to sustain when they have positive outcome expectations which leads to success. Moreover, instructional factors such as effective scaffolding, proper instruction, or feedback may promote students’ engagement and further promote self-efficacy (Schunk, 2008).

Computer Simulations

Computer simulations are widely recognized as educational tools to promote and facilitate student learning. With the increasing focus on integrating technology into education, computer simulations have been extensively studied. As indicated by Baek (2009), simulations can support constructivist learning by allowing students to learn by
doing in a resource-rich environment. According to Nickerson (1995), simulations can not only draw students' attention to aspects of problems that can be easily missed, but also create an atmosphere in which ideas can be freely expressed and provide encouragement when students makes an effort to understand. In science, simulations can reduce the cost of expensive laboratory equipment and protect students from dangerous experiments (Blake & Scanlon, 2007). Furthermore, if simulations are well-designed, they can save instructors a lot of time on explanation and provide more opportunities to interact with students.

Of all the advantages, the most important characteristic of computer simulations is offering students problem manipulation space. According to Jonassen (1999), in order to keep students actively engaged in learning activities, they have to manipulate something and see the effects of their manipulation since “students cannot assume any ownership of the problem unless they know that they can affect the problem situation in some meaningful way” (p. 222).

Reigeluth and Schwartz (1989) identified three major types of simulations. The first type is a procedural simulation, which teaches learners how to perform a sequence of steps or decisions, such as driving a car or solving an equation. The second type, a process simulation, refers to those that teach naturally occurring phenomena composed of a specific sequence of events, such as the action of a volcano or the change of four seasons. The last type is a causal simulation, which teaches the principle of causes and effects, such as central limit theorem or the law of supply and demand.

Based on an investigation of instructional theories and simulations, Reigeluth and Schwartz (1989) further identified five essential characteristics in a simulation:
generality, example, practice, feedback, and help. Generality refers to either visual or auditory presentation of the targeted content. An example is a demonstration of the relationship among changes described in one or more generalities. Practice allows learners to apply one or more generalities to diverse situations. Feedback gives learners positive or corrective information regarding their responses. Help provides learners with assistance when they have a problem during generality, examples, practice, and feedback phases.

Studies within a variety of disciplines acknowledge the effectiveness of computer simulations. Tsai et al. (2008) developed a computer simulation using virtual reality (VR) to teach novice nurses how to perform Port-A cath injection. Post-test scores were significantly higher among the VR group than the control group. In Ronen and Eliahu’s study (2001) of a simulation to teach electric circuits, significant differences were found between the achievement scores of students who performed the tasks with and without the simulation. After evaluating three online simulations, Blake and Scanlon (2007) pointed out that the use of simulations could enhance students’ acquisition of scientific concepts.

In statistics education, many researchers have proposed using computer simulations as an effective tool to promote understanding.

Computer Simulations for Statistics Education

Researchers and educators in statistics education often complain that students’ misconceptions and faulty intuitions are resilient and difficult to change (Garfield, 1995; Garfield & Ben-Zvi, 2007). This is particularly evident with abstract concepts, such as predicting the results of a random process that was repeated indefinitely, which are
difficult for most people to grasp. Thus, statistics teachers are always looking for approaches to improve statistical instruction and enhance students’ statistical reasoning.

The computer simulation method (CSM) is one approach that researchers and teachers recommend (delMas et al., 1999; Hsu, 2003; Mills, 2000; Yu, Behrens, & Anthony, 1995). Using CSM provides opportunities for students to learn by interacting with the program and observing the process, which makes abstract concepts more concrete (Garfield & Ben-Zvi, 2007). A review of the literature found few studies that examined the effectiveness of CSM on statistical learning. Those that were found included comparisons of different approaches, such as using CSMs and textbooks (Lane & Tang, 2000; Mills, 2000), pretest and posttest (Garfield, delMas, & Chance, 1997), and students’ attitudes toward using computer simulations.

Mills (2000) investigated the effects of using CSMs as a teaching tool to enhance introductory students’ understanding of abstract concepts. She conducted two studies: teaching confidence intervals to 212 undergraduates in an introductory statistics course, and teaching Central Limit Theorem (CLT) to 32 graduate students in an introductory statistics course. Those in the CSM groups executed Excel commands to illustrate abstract concepts, such as performing Excel commands to generate a large number of samples from the uniform population, to calculate the mean from each sample, and to plot it in the histogram. Explanation of the effects of different sample sizes and population distributions for sampling means was provided. Students were able to personally examine the effects by changing the sample size and population distributions.

Using scores in the mathematic section of the Scholastic Aptitude Test (SAT) and the quantitative section of the Graduate Record Examination (GRE) as a covariate for
both studies, results showed that there was no statistically significant difference between CSMs and traditional approach groups on improving students’ understanding of abstract concepts. The researcher indicated that the insignificant findings may partly be due to the posttest items with low reliability estimates (Cronbach Alpha), which implies that the items were too difficult for introductory students to understand. Another reason may have been the instructional unit issue. As Mill pointed out, generating samples from a population with known parameters in Excel was time-consuming and executing the commands distracts learners’ attention from the important concepts.

Findings from Lane and Tang’s research (2000) suggested that students trained by computer simulations significantly outperform those trained with a textbook on acquisition of sampling distributions. A 2 by 2 factorial design including medium (simulation versus textbook) and question specificity (specific versus non-specific) was employed, together with a control group with no treatment. Subjects were 115 undergraduate students with no formal statistical training who took a 12-minute Wonderlic Personnel Test as a general measure of cognitive ability. The authors used the test scores as a covariate and the computer simulation was pre-defined (video clips). Subjects spent 30 minutes watching the simulation projected on a big screen while listening to the narration. They were unable to explore or interact with the simulation on their own. Results showed that subjects trained by the simulation display performed significantly better than those trained with a textbook. There was no statistically significant difference between specific and non-specific question conditions. The same experimental design was conducted again in 2002. Lane and Tang (2002) added a second experiment replacing the question specificity factor with the timing of posttest. The
findings indicated that subjects trained with the simulation display scored higher than those trained with the textbook on the transfer test about statistical reasoning.

In another study, Garfield et al. (1997) had students use a computer simulation to learn sampling distributions. Participants \( n = 101 \) were asked to change parameters (e.g., the population shape and sample size) and to summarize results of the different sampling distributions they observed. Comparing scores of the pre- and posttest, the researchers found that although there was a significant improvement from pretest to posttest, many students still had difficulty understanding sampling distributions after the treatment. To solve this problem, Garfield et al. conducted a second study and redesigned an activity employing the conceptual change theory to engage students in identifying their misconceptions and to help them solve faulty intuitions. Results indicated that the new activity was an effective approach with a statistically significant improvement in posttest scores. The researchers indicated that learners’ prerequisite knowledge plays an important role in the program since sampling distributions combine many concepts learned from previous courses.

Similarly, Jones, Hagtvedt, and Jones’ study (2004) suggested the importance of prerequisite knowledge. They developed a computer simulation using VBA (Visual Basic for Application) in Microsoft Excel to facilitate students’ acquisition of regression concepts. The authors found that learners’ prerequisite knowledge of the learning content and encouragement for students to spend more time on the tutorials were essential. Online survey results indicated that students had a positive attitude toward using this software to learn statistical concepts, which agrees with many researchers’ opinions
(Aberson, Berger, Healy, & Romero, 2002; Jones et al., 2004; Lane & Tang, 2002; Mills, 2000; Hagtvedt, Jones, & Jones, 2008).

In sum, there is no specific method for using computer simulations to teach statistics. Some instructors prefer spreadsheets, some favor Java applets, and still others play video clips of simulations. A review of literature provides valuable ideas for current research, such as approaches to help students recognize their misconceptions and then providing instruction. Watching a simulation is similar to viewing a worked example with step-by-step guidance as a form of learning support. Embedding worked examples into a computer simulation can be an effective approach for transfer of learning with beginners (Meier, 2007). Finally, providing the prerequisite knowledge for the learning content and encouragement for students to spend more time on the program should be taken into consideration as well.

Most simulations focus on illustrating concepts and solving problems. Very few of them provide students with scenarios and allow them to see how these concepts are used in real life. Although a few studies did provide scenarios, what they provided was limited to a short paragraph of text description. The next section presents an instructional design model, goal-based scenarios.

Characteristics of Goal-Based Scenarios

A goal-based scenario (GBS) is a computer-based simulation of learning-by-doing (Schank et al., 1993/1994). It is an instructional design model based on learning theories such as constructivism, situated cognition, and conceptual changes (Reigeluth, 1999). In a GBS, students can work individually or in small groups to play a role in a problem scenario. They are responsible for pursuing a goal by practicing skills or
gathering and applying relevant information to solve problems (Schank et al., 1999). These problems, selected by the designer to attract students' interests, are moderately structured (Lohman, 2002); learners must work with the content and paths that are specified by the simulation, even though they are allowed to take a variety of paths to gather information and achieve their goal. During the simulation, instruction, worked examples, well-told stories by experts, or other resources are given to learners if they have no idea of how to perform the task or are curious about something (Schank et al., 1993/1994). In addition, video clips are provided as feedback to students’ work. A GBS ends when a learner completes the task specified by the simulation.

Basically, what a GBS teaches are skills and complex systems (Schank et al., 1993/1994). Skills mean that people know how to do something to achieve their goals. The complex systems are the contexts to which the skills are applied. The purpose of a GBS is to teach skills in a context that is simulated to present a real-life environment in order to help students index relevant information, make predictions, and create explanations for the various phenomena taking place around them (Brown et al., 1989; Schank et al., 1999). Teaching without context degrades instruction to teaching facts, which may not exist long in students’ mind and may difficult to apply (Schank & Cleary, 1995). A GBS model consists of seven components: learning goals, mission, cover story, role, scenario operations, resources, and feedback (Schank et al., 1999).

**Learning Goals**

Learning goals are target skills that a course designer or instructional designer wants students to learn. They have two different categories: process knowledge and
content knowledge. The former is knowledge about how to practice skills that helps one achieve the goals. The latter is the information needed to attain the goals.

**Mission**

The mission is an interesting, realistic, and motivational task for the students to pursue. Students must use the target skills and knowledge to complete the mission so that they can achieve the goal successfully.

**Cover Story**

The cover story is the background story that creates the need for the mission to be completed and offers learners sufficient opportunities to search information or practice skills. As with the learning goal and mission, the cover story should be interesting, motivational, and somewhat realistic.

**Role**

The role is the character the user plays in the cover story. It should be a role that is appropriate to practice the necessary skills or use the knowledge in the scenario. Like the goal, mission, and cover story, the role should be realistic and motivating. Role-playing has been a recognized and effective approach to engage students in their learning experience (Lowenstein, 2007). A GBS model incorporates this element to foster learners’ motivation (Dickey, 2007). The use of role play, according to Van Ments (1999), is “asking someone to imagine that they are either themselves or another person in a particular situation. They are then asked to behave exactly as they feel that person would” (p. 4-5). By doing so, students are provided with an opportunity to investigate why people perform as they do, to identify the systematic steps in decision-making or
problem-solving process, and to become actively involved in more realistic learning experience (Lowenstein, 2007).

**Scenario Operations**

The scenario operations are all the activities that students do in the GBS in order to complete the mission. They can be anything that the designers or educators think will promote students’ comprehension. To learn statistics, the scenario operations could be asking for expert opinions, running simulations to visualize the concept, or reading relevant information. Students can select their favorite operation(s). Within the scenario, little time is spent on explaining the scenario in detail and more time is spent on practicing skills and acquiring knowledge. The scenario operations also include decision points together with positive or negative consequences as reinforcement.

**Resources**

Resources provide information that students need in order to acquire the target skills or content knowledge to complete the mission successfully. Resources in a GBS have two types. The first type is well-organized information such as text, video clips, narration, graphics, or other materials that are accessible to students. The second type of resource is stories that are embedded with lessons. Coming from case-based reasoning theory, Schank and Cleary (1995) believed that people can learn from other people’s stories or cases. Stories make events more memorable by offering many indices such as contexts, decisions, time, characters, or something that attracts people’s attention (Schank, 1990).
Feedback

Feedback is given in three ways. The first is through the consequence of action. An example, when a learner makes a mistake in a GBS, a video clip will play with a client’s complaint. The second feedback mechanism is instant feedback from an online instructor that monitors students’ learning progress. The final feedback method is through experts’ stories. Students learn from experts’ stories about how they deal with a problems or events.

Human-led GBS is also called a live GBS. With instructor coaching, students learn skills in the scenario through real human interaction or group interactions (Schank et al., 1999). Teachers can interrupt the process to give instruction in real time. This instant feedback, including real interaction with mentors and an unlimited and various set of responses, is what a computer-based GBS lacks (Schank et al., 1999; Macpherson, 2006). Conversely, advantages of a computer-based GBS include the opportunity for learners to work alone, to fail without embarrassment in front of peers, to repeat using the same program anytime and anywhere until they acquire the skills, and to receive instruction from world-class experts through video clips (Schank, 2002). The next section presents current research on goal-based scenarios.

Current Research on Goal-Based Scenarios

Most studies about GBS focus on introducing how computer simulations were developed and how learners interacted with the programs, such as Innmasters by Campbell and Monson (1994) to train hotel management and Sickle Cell Counselor by Bell, Bareiss, and Beckwith (1993/1994) to instruct museum visitors about sickle cell
disease. Few studies had been conducted to investigate the effectiveness of GBS on learning outcomes.

Within existing limited studies, Schoenfeld-Tacher, Jones, and Persichitte (2001) found that the posttest scores of participants (458) improved significantly with use of GBS. They developed a science lesson using a GBS model to teach college students about DNA. The simulation, called Whodunnit?, was made available to students through WebCT (a web-based classroom content management system). Participants played the role of an expert to assist local authorities in solving a murder case. The instructional unit included the activities of The Crime, The Suspects, GBI Academy, Internship, Virtual Crime Lab, and Courtroom. After the pretest, participants had four days to complete the simulation including the instruction, lab report, and reaction questionnaire. They took the posttest in class on the fifth day. Taking the treatment for four days is acceptable since Schank and Cleary (1995) suggested that learners in a GBS learning environment should have sufficient time to construct their own knowledge and achieve the learning goal. The findings indicated that the learning outcomes had no relationships with gender or ethnicity, except for scores on the pretest and the logical thinking test.

Some research results indicate that students have positive attitudes toward using GBS programs (Mouza & Bell, 2001; Naidu et al., 2000). Mouza and Bell (2001) developed an online science learning environment, called ALPINE, to help primary and middle-grade students acquire weather concepts. Students worked in teams to collect information, interpret data, and draw conclusions in order to decide the most appropriate training location for the US ski team. They could use forums to communicate with teammates or to compare their findings with those of the other teams. They were also
provided a wide array of Internet resources to promote in-depth exploration of other
weather sites. Surveys indicated that teachers and students had a positive attitude toward
using ALPINE in the classroom. Mouza and Bell found that without appropriate training
and understanding of the program, teachers could inhibit technology use in the classroom.
Some students complained that they were restricted to use a certain functions in the
program, which failed to solve their inquiry through multiple solution paths.

In sum, a sound computer simulation needs an appropriate instructional design
model that offers explicit guidance on how to better help people learn and develop. After
all, simulations do not work on their own; without approaches to promote students’
interaction with computer simulations, a simulation is unlikely to be an effective teaching
tool (Blake & Scanlon, 2007). With the increasing trend of using computer simulations to
teach statistical concepts, there is a need to apply an instructional design model that has
been verified and tested to develop computer simulations.

Problem-based learning (PBL), case-based learning (CBL), and goal-based
scenario (GBS) share similar characteristics in nature. The next section explains the
distinctions among these three approaches.

Distinction among PBL, CBL, and GBS

Many researchers agree that learning driven by problem solving may enhance
motivation, integrate knowledge and practice, and increase the visibility of students’
reasoning processes (Barrows & Tamblyn, 1980; Hmelo-Silver, 2004; Thomas,
O’Connor, Albert, & Brandt, 2001; Williams, 2005). Problem-based approaches are a big
family including problem-based learning (PBL), case-based learning (CBL), and goal-
based scenario (Hmelo-Silver, 2004; Macpherson, 2006; Williams, 2005). Although these methods share similar features, they still differ in some aspects.

PBL is a learner-centered method that has its origins in medical education (Hmelo-Silver, 2004; Hmelo-Silver et al., 2007). In PBL, learning occurs when students are working collaboratively toward solving realistic problems (Barrows & Tamblyn, 1980). These problems are ill-structured and have unstated goals and constraints that require them to make judgments about the problems depending on their personal opinions or beliefs (Jonassen, 1999). Norman and Schmidt (1992) contended that by using routinely occurring problems with many critical features, students will transfer concepts from the prototype problem to clinical practice. Students not only learn content and strategies but also self-directed learning skills (Hmelo-Silver, 2004). Through collaboratively solving problems and reflecting on their experiences, learners are cognitively engaged in sense-making, developing explanations, and communicating their ideas (Hmelo-Silver et al., 2007).

Researchers are in agreement (Barrows & Tamblyn, 1980; Hmelo-Silver et al., 2007; Jonassen, 1999) that PBL offers many forms of scaffolding to support students’ zone of proximal development (Vygotsky, 1978). An instructor becomes a facilitator who provides necessary resources and guides students through the learning activity. Unlike the subject-based learning that uses problems as an exercise for applying the information already learned, students in PBL encounter a problem first, which may draw learner’s attention or serve as a stimulus for practicing problem-solving or reasoning skills (Barrows & Tamblyn, 1980; Hmelo-Silver, 2004; Jonassen, 1999). They formulate and analyze the problem by identify facts in the scenario. While striving to understand the
problem better, they generate hypotheses and recognize knowledge deficiencies that later become learning issues during their self-directing learning. They then apply their new knowledge, evaluate their hypotheses, and reflect on when they have learned.

Case-based learning is a specialized type of discussion that requires students to either individually or in groups perform analysis and seek a solution to the problem presented by the case (Carlson & Schodt, 1995). Similar with PBL, the purpose of CBL is to enable students to learn by doing and promote fruitful discussion (Flynn & Klein, 2001). Students in CBL are introduced to the objectives first and then engage in preparing the fundamental information about the case, determine concepts or principles applying to the case, investigate possible solution and causes, select appropriate solutions, and discuss solution implementation (Lohman, 2002).

CBL tends to be expert-oriented because a teacher not only identifies important facts and topics appropriate for discussion, but also directs the discussion, produces intermediate summaries, controls the flow of discussion, and wraps up the discussion. Moreover, the teacher supports students with pertinent readings that deal with different aspects of the case in order to draw their attention to critical ideas (Harling & Akridge, 1998). Since information needed for the case is organized and offered in CBL, students become less responsible for problem framing and the problems they solve tend to be well-structured.

Goal-based scenarios can utilize either a human-led or computer-based approach (Schank et al., 1999). Since the design of each approach is different and most studies have been conducted by employing the latter, a computer-based GBS was used to compare with PBL and CBL.
In GBS, learners work individually or in a small group within a computer-simulated environment to accomplish a goal and learning specific problem-solving skills (Schank et al., 1999). They are first presented a goal and mission and then try to achieve the goal while playing a role in the scenario. Although they are allowed to take a variety of paths to gather relevant information, they must work with the information and paths that are specified by the simulation since these resources (such as problems, computer-based contents, and feedback) are pre-designed (Lohman, 2002). Thus, a GBS tends to foster problem-solving skill in well-structured domains.

All three approaches employ a role-playing method to engage learners in solving authentic problems (Carlson & Schodt, 1995; Hmelo-Silver et al., 2007; Schank et al., 1999). Resources in PBL are flexible since self-directed learners individually seek information for their different inquiries (Hmelo-Silver, 2004). Resources in CBL are somewhat flexible because instructors prepare and direct learning activities and can slightly modify the materials as needed (Carlson & Schodt, 1995,). In GBS, learning content is programmed in the computer-based tutorial, so resources are less flexible than PBL and CBL.

CBL and GBS are likely to promote single-loop learning that “occurs when errors are detected and corrected without altering the governing values of the master program” (Argyris, 2005, p. 262). This occurs because students in CBL and GBS are provided with well-structured problems and resources. In addition, the features of these two approaches limit their engagement in problem framing (Bardwell, 1991), identify solution procedures for them, and restrict the range of acceptable solution(s) (Lohman, 2002). In contrast, learners in PBL tend to solve ill-structured problem, which results in double-loop
learning that “occurs when, in order to correct an error, it is necessary to alter the governing values of the master programs” (Argyris, 2005, p. 263).

Although sharing similar features, there still exist differences between the design of games and simulations. The next section will identify these distinctions.

**Distinctions between Simulations and Games Design Model**

There is a trend towards shifting from a traditional and didactic model of instruction to a learner-centered model through the application of computer games and simulations (Garris, Ahlers, & Driskell, 2002). Not only are students provided with unique opportunities to interact with a knowledge domain, but also with an engaging environment for learning (Kiili, 2005). Although games and simulations (in which a GBS model belongs to) share similar features, they still have some distinctive aspects.

To examine the differences between a simulation and a game design model, some salient features of game design must be identified first. However, researchers’ opinions regarding necessary game components differ greatly. For example, Prensky (2001) suggested that digital games must consist of six structural elements: 1) rules, 2) goals and objectives, 3) outcomes and feedback, 4) conflict, competition and challenge, or opposition, 5) interaction, and 6) representation or story. Sauvé, Renaud, Kaufman, and Marquis (2007) stated that the essential elements are player(s), conflict, rules, predetermined goal of the game, and its artificial nature. Reviewing more than 20 studies, Garris et al. (2002) proposed six main game characteristics, including fantasy, rules and goals, sensory stimuli, challenge, mystery, and control. Among these features the researcher selected fantasy, rule and goal, and challenge to investigate the distinctions between simulation and game design models.
**Fantasy**

Games are equated with having fun. In order to provide an exciting and emotionally appealing experience, such as exploring in an alien world or fighting a beast, games often deviate from reality (Prensky & Thiagarajan, 2007). Researchers (Garris et al., 2002; Prensky & Thiagarajan, 2007) indicated that fantasy is encouraged by technology and its creation not only provides plenty of opportunities for users to interact in imaginative situations, but also facilitates the focus of attention. In addition, the emotional appeal of fantasy generates cognitive and sensory curiosity that promotes motivation in digital games (Habgood, Ainsworth, & Benford, 2005). Thus, games do not have to represent reality.

In contrast to games, simulations tend to be based on reality (Hertel & Millis, 2002). Simulations cannot use imaginative elements as a representation of the real world. It is because they are designed to replace activities aimed at experiencing and investigating real phenomena and concepts in order to bridge the gap between theories and students’ understanding (Ronen & Eliahu, 2001). Later they can accurately apply what they learn to the real-life context after completing the tutorial. Thus, simulations have to accurately represent reality in the simulated context and games do not.

**Rules and Goals**

Both games and simulations consist of two essential components: rules and goals. The former imposes limits and forces players to take specific paths to reach goals in the game world, whereas the latter is an important factor of triggering their motivation and attention (Prensky & Thiagarajan, 2007). Goals in games tend to encourage players to competitively win the objective of games, such as getting the highest score, while
simulations encourage users to accomplish specific tasks with competition of little importance (Akilli, 2007). Goals in games can be any interesting goal to entertain players and acquisition of knowledge is unnecessary, while students in simulations must execute serious responsibilities with associated consequences (Gredler, 1996). Rules in games can be imaginative and not relate to the real world, whereas a simulation is to investigate “dynamic set(s) of relationships among several variables that may change over time and reflect authentic causal processes” (Gredler, 1996, p. 523). Thus, the all processes or interaction should result in predictable ways.

**Challenge**

Challenges in games or simulations refer to human or computer-controlled obstacles that prevent players from achieving the goals easily (Narayanasamy, Wong, Fung, & Rai, 2006). As researchers (Garris et al., 2002; Rieber, 1996) indicated, appropriate and progressive task difficulty may promote learners’ motivation and enjoyment. Challenges are utilized differently in games and simulation. Take a time limit for instance. In games, challenges are often used by imposing a time limit to promote competition and excitement (Garris et al, 2002), whereas simulations do not set up a time limit since learners are encouraged to try different strategies and see the results (Blake & Scanlon, 2007). Moreover, each learner may need different amount of time to construct their knowledge due to individual differences. Challenges in games are combined with luck since the factor of luck gives players a chance to win and boost their motivation (Hertel & Millis, 2002). In contract, participants in simulations typically have more predictable consequences of their actions while dealing with challenges.
Despite the distinctions between games and simulations, they are correlated to each other (Aldrich, 2005; Prensky, 2001). Instead of considering the distinction between simulations and games, Aldrich (2005) proposed that it is more productive to think about how simulations, games, and pedagogy relate to each other. A good use of these three elements can contribute to appropriate educational experiences, which was called an education simulation by Aldrich (2005) or a game-based training by Martens, Diener, and Malo (2008).

Simulation, as defined by Aldrich (2005), is to “selectively represent objects or situations, and selectively represent user interaction… [in order to] enable discovery, experimentation, role modeling, practice, and active construction of [content]” (p. 81). Its ultimate goal is to allow students to transfer what they have acquired in the simulation to the real world (Aldrich, 2005; Schank et al., 1999). Game elements are highly interactive experiences providing entertainment (Prensky, 2001). Pedagogical guidelines describe how to teach and train in order to prevent superficial behaviors (Aldrich, 2005).

Based on Aldrich’s framework, Martens et al. (2008) further described the overlap of these three components. Leaving out the simulation aspects may create edutainment games, whereas leaving out pedagogy aspects may create simulation games. Finally, leaving out game may create a training simulation, which agrees with Prensky and Thiagarajan’ (2007) statement that “simulations are not, in and of themselves, games. They need all the additional structural elements—fun, play, rules, a goal, winning, competition, etc.” (p. 212). In this study, training simulation is the category a GBS simulation belongs to.
Before distributing instructional materials to the target population, formative evaluation is an essential, ongoing step that is conducted to avoid problems and improve instructional effectiveness. The following section will introduce ways to perform formative evaluation.

Formative Evaluation

Formative evaluation is an iterative process that a designer utilizes to collect users’ data for analysis and revision in order to make instructional materials more efficient and effective (Dick, Carey, & Carey, 2005). Its purpose is to identify specific errors in the materials, such as instrument and procedure, and correct them. For this study, a qualitative methodology called formative research was adopted as an evaluation method to improve the design theory or model (English & Reigeluth, 1996).

Formative evaluation consists of three stages (Dick et al., 2005): one-to-one evaluation, small-group evaluation, and field trial. The initial phase allows a designer to work individually with target learners and revise errors in the materials. In the second stage, the instructional materials are tested by a small group of students who work on the materials alone. The desired information, such as testing the revisions from the first phase, is collected during their trials. The final stage is a field trial that examines how close the implementation of instructional materials matches a situation in real context.

There are three criteria for formative evaluation (Dick et al., 2005): clarity of instruction, impact on learners, and feasibility. To understand clarity of instruction, a designer may ask questions, such as how clear is basic information presented to the learner? How are the sequence, the size of section, the pace, and the transition tailored for the learner? The second criterion, impact on learners, refers to learners’ attitudes about
the instruction and achievement. Questions an evaluator may ask include: 1) Is the instruction personally relevant to learners?, 2) Is the instruction accomplishable with reasonable effort?, 3) Is the instruction interesting and satisfying to experience?, and 4) Will the achievement tests examine whether the individual can recall the information and perform the tasks? Finally, the criterion for feasibility refers to the capability of the learner and the instructional medium. Questions of interest include: 1) “How will the maturity, independence, and motivation of the learner influence the general amount of time required to complete the instruction? (p. 285),” and 2) is the learner comfortable with instructional medium?

When learners’ feedback is collected, the next critical step is to analyze and interpret the data. Dick et al. (2005) suggested that interpretation of data gathered from one-on-one trials should not be overgeneralized since there is no guarantee that the next user will come up with similar responses. Designers can immediately correct frequent errors and leave questionable ones until the instructional materials are retried by different users or a small group (Dick et al.).

In summary, formative evaluation of instructional materials is conducted to enhance efficiency and effectiveness of the program. In this study, the researcher merely conducted one-on-one evaluation on the instructional program, due to its focus on investigating an instructional design model. In addition, while evaluating the design instance, the researcher took those three criteria including clarity of instruction, impact on learners, and feasibility into concern, and used several trials to verify questionable feedback.
Understanding how learners develop a concept is essential because it helps to prescribe precise interventions or activities that promote students’ reasoning. The use of a developmental model of sampling distribution is to serve this purpose and it is introduced in the following section.

Developmental Model of Sampling Distributions

Over the past decade, practitioners in statistical education have urged more emphasis on statistical reasoning (Garfield & Ben-Zvi, 2007). Statistical reasoning can be defined as “the way people reason with statistical ideas and make sense of statistical information” (Garfield & Ben-Zvi, 2007, p. 381). This includes interpreting data or graphs, using a summary to make predictions, or reasoning about samples and population (Garfield, 1998; Ben-Zvi, & Garfield, 2004). Students who develop mature statistical reasoning may connect one concept to another, be able to explain processes, and be able to interpret statistical results. Statistical reasoning is often taught through direct instruction in which instructors often stress skills, procedures, and computations (Chance et al., 2004). This approach may be too abstract for students to comprehend and does not lead them to reason or think statistically (Ben-Zvi & Garfield, 2004; Fong, Krants, & Nisbett, 1993).

Using a cognitive model of development is an effective approach to improve students’ statistical reasoning (Jones, Langrall, Mooney, & Thornton, 2004). According to Jones et al. (2004), the model is formulated through structured interviews and teaching experiments that enable researchers and teachers to trace how students’ statistical reasoning is built up during instruction. An advantage of using a developmental model is that it may help teachers prescribe more precise interventions or activities to enhance
students’ statistical reasoning by accurately placing students in different dimensions of the model (Chance et al., 2004). Further, the model can help instructors evaluate and monitor students’ performances over time. Students are more motivated and receptive to instruction while the model helps them recognize their misconceptions (Chance et al.; Garfield, 1995).

Based on Jones, Langrall, Mooney, and Thornton’s developmental model of statistical reasoning in children, a framework was developed by Chance et al. (2004) to describe developmental stages of students’ statistical reasoning about sampling distributions. These are: 1) idiosyncratic reasoning, 2) verbal reasoning, 3) transitional reasoning, 4) procedural reasoning, and 5) integrated process reasoning.

Learners at the phase of idiosyncratic reasoning may understand relevant information about sampling distributions, but they do not know how to use it or may use it without fully understanding it. When asked to explain the concept, they may come up with irrelevant information.

The second phase is verbal reasoning. Learners at this stage are able to recite the definition and explain the importance of sampling distributions, but when it comes to critical conceptions, they still have difficulty in understanding relationships or how to apply the concepts.

The third stage is transitional reasoning. Students start constructing the concepts based on some, but not all, of the concept’s attributes. For example, they might be able to identify one or two key attributes of sampling distributions that consist of three critical ideas: center, spread, and shape. If the sampling is random and the sample size is large, center, the mean of sampling distributions, will be centered at the mean of the population.
The spread of sampling distributions will tend to be less spread-out and the shape of sampling distributions will be close to normal.

The fourth stage is procedure reasoning. Learners still have difficulty putting these concepts together even though they can recognize the three key concepts of sampling distributions. For instance, they cannot confidently predict what the shape or spread of sampling distributions will be when given some parameters about the population or a sample size.

The final stage is integrated process reasoning. At this level, students develop a thorough comprehension about sampling distributions. They are able to confidently and accurately describe the concepts in their own words, explain the relationship among key points, and predict the process of sampling distributions. For example, when given an abnormal population with the standard deviation and mean, students can explain what characteristics of sampling distributions will be according to a given sample size.

Gender Gap in Technology Use

The ubiquity of computer usage in educational context has grown significantly. Although educators, researchers, and students have an optimistic perception of using technology in the classroom (Plumm, 2008), gender gap in computer use still exists among students. As indicated by Plumm (2008), due to inexperience with technology and social stereotypes that exclude women from math, science, or technology-related domains, female students tend to have less positive attitudes toward computer use. In Pierce, Stacey, and Barkatsas’ study (2007) investigating students’ affective attitudes toward learning mathematic with technology, results showed that only male students’ attitudes were correlated with confidence in using technology. Similarly, results of
Sieverding and Koch’s research (2009) examining undergraduate students’ computer competence indicated that females tended to judge their computer competence to be lower than similar self-assessments by males. Researchers (Sieverding & Koch, 2009; Vekiri & Chronaki, 2008) found that women attributed the failure to their own inability and men attributed the deficiency to external causes, such as malfunction of equipment. This unfavorable attribution, according to Koch, Muller, and Sieverding (2008), may further influence students’ computer self-efficacy or learned helplessness.

To narrow the gender gap in technology use, Cooper (2007) offered three suggestions. First, instructional designers or software developers should alter gender stereotype by designing programs that appeal to females as much as males. For instance, possible approaches may include focusing on gender equity as role models in the tutorial rather than presenting females as secondary roles or as helpless and in need for rescue (Canada & Brusca, 1991; Cooper, 2007). Second, schools and teachers should provide appropriate support for females to interact with computers such as offering private or small same-sex groupings, since mixed gender groups may discourage females. Finally, instructors and parents should be aware of females’ attributions and help guide their causes more externally for computer glitches and more internally for their successes and efforts.

Summary

The review of literature on the theoretical foundation of learning supports the idea that learning becomes effective when students are engaged in participation of practice and their own knowledge construction. Instruction should promote the generation of disequilibrium, encourage students’ exploration, and provide sufficient scaffolding. In the
aspect of motivation, variables that influence learning consist of goal-setting, self-efficacy, peers, context setting, pedagogy, and attribution. These variables are cyclic and each one will affect future learning and motivation. A review of studies that used CSMs to illustrate statistics concepts provided valuable ideas for this study, such as engaging students to recognize their misconceptions first and providing instruction after, using worked examples, providing prerequisite knowledge, and encouraging students to spend more time on the program.

To better understand GBS, a detailed review of its components and relevant research were provided together with an investigation of distinctions among PBL, CBL, and GBS, and differences between GBS and game design model. Finally, guidelines of formative evaluation and the developmental model of sampling distribution were reviewed to facilitate the design of a sound instructional instance.
CHAPTER 3: METHODOLOGY

Introduction

The purpose of this study was to evaluate and explore possible improvements in an existing instructional model, goal-based scenarios (GBS). This model was proposed more than 15 years ago and was slightly modified by Schank, Berman, and Macpherson (1999). Although many simulations were developed using GBS, no empirical studies have investigated strengths, weaknesses, or possible improvement of this model. Thus, the focus of the present study was to examine GBS and provide insights into the model and its application. The research questions were: 1) What are the strengths and weaknesses of the GBS model?, and 2) What improvements can be made?

To answer these questions, it is most appropriate to apply qualitative methodology since student feedback on instructional improvement cannot be easily acquired through a quantitative approach (Shon, 1996). As Reigeluth and Frick (1999) indicated, “traditional quantitative research methods (e.g. experiments, surveys, correlational analyses) are not particularly useful for improving instructional-design theory, especially in the early stages of development” (p. 634).

Formative Research Methodology

A qualitative approach to evaluate and improve an instructional theory (or model) is formative research. Proposed by Reigeluth and Frick (1999), its core tenet is that “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” (Reigeluth & Frick, 1999, p. 636). Thus, questions guiding formative research are: what
approaches work well, what approaches do not work well, and what improvements can be made to the model?

Formative research is similar to formative evaluation, which refers to a process of refinement that designers often utilize to make a program more efficient and effective (Dick et al., 2005). The primary difference is that the purpose of formative research is extending and improving the knowledge base about instruction (models and theories), rather than improving an instructional program or system (English & Reigeluth, 1996). To do that, an instance that is based as faithfully as possible on guidelines from an instructional design model or theory is created or identified first.

The next step of the formative research process is to conduct data collection and data analysis. Data collection is a series of one-on-one formative evaluations of the design instance with learners from the targeted population. They are encouraged to identify strengths and weaknesses of the instance and suggest ways to improve it since any weakness in the instance may reflect ways to improve the theory (English & Reigeluth, 1996). A case study approach is adopted in formative research to collect data, such as using interview to probe the thinking of the participants and applying member checking to check validity of the data (Reigeluth & Frick, 1999). In addition, data analysis is conducted during the process of data collection not only to identify strengths, weaknesses, and possible improvements, but also to check the consistency of data across participants.

When data from the first round of data collection and analysis is completed, the next step is to revise the design instance in terms of these data. Following this step, the data collection and revision cycle is repeated until findings reach saturation. The
researcher can offer tentative revisions for the theory, but these suggestions require future studies to thoroughly replicate and validate these revisions before they become “knowledge” (Reigeluth & Frick, 1999). Studies that used the formative research method to improve instructional design models or theories include the elaboration theory (Kim, 1994), a theory for designing computer-based simulations (Shon, 1996), a model for facilitating systemic change in public school districts (Joseph, 2003), a model for designing virtual reality based learning environments (Chen, 2007), and a theory for designing educational video games (Watson, 2007).

The present study employed formative research using a designed case. According to Reigeluth and Frick (1999), formative research cases can be classified as designed or naturalistic. In the study of designed cases, a researcher develops a tutorial based on an instructional model or theory and formatively evaluates the instance. In the study of naturalistic cases, a researcher would choose an instance that was not specifically designed according to the model, but shared the same purpose with the model, analyzes how the instance relates to the model, and formatively evaluates the instance to identify how each component could be refined. Since a computer simulation based on the GBSs model was developed with the intent to formatively evaluate the tutorial, this study would be classified as a designed case.

The procedure of conducting formative research consists of six steps (Reigeluth & Frick, 1999):

1. Select a design theory.
2. Design an instance of the theory.
3. Collect and analyze formative data on the instance.
4. Revise the instance.
5. Repeat the data collection and revision cycle.
6. Offer tentative revisions for the theory.

The implementation of each step in the present study is explained in the next section.

Formative Research Study Design

Select a Design Theory

The focus of formative research, as Reigeluth and Frick (1999) proposed, is to improve an existing instructional design model (or theory) and the first step is selecting a model. In the present study, the researcher wanted to evaluate and improve the GBS instructional design model. Refer to Chapter 2 for a detailed review of GBS. Although GBS was proposed more than 15 years ago, no empirical study has been conducted to investigate this model. The goal of this study was to examine and improve the guidance GBS offers.

Design an Instance of the Theory

An instructional program named Statistics Specialist was designed by the researcher to teach students sampling distributions and evaluate the application of the GBS model. Reigeluth and Frick (1999) emphasized that “the design instance should be as pure an instance of the design model as possible” (p. 639). They suggested that researchers truly include components of the model and avoid those that are not called by the model. This is related to construct validity. Table 1 indicates how Statistics Specialist was designed based on the framework of the GBSs model. As seen in Table 1, scenario operations consist of two approaches: expository and discovery. The former refers to the
function of asking the expert; the learner asks the expert questions and the expert offers explanation. The latter refers to running the simulation; the learner follows the guideline to perform a simulation so that they can figure out how the target ideas develop.

### Table 1

**Framework of Statistics Specialist**

<table>
<thead>
<tr>
<th>Features of GBS</th>
<th>Statistics Specialist</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Goal</strong></td>
<td>Teaching students the concept of sampling distributions.</td>
</tr>
<tr>
<td><strong>Mission</strong></td>
<td>Accurately advising the client, Ken.</td>
</tr>
<tr>
<td><strong>Cover story</strong></td>
<td>Ken is a new employee working on a shrimp farm. His boss is currently trying a food supplement to accelerate the growth of the shrimp. In order to see how this supplement works, the boss wants him to monitor the growth of the shrimp, particularly, measuring the average weight of shrimp weekly.</td>
</tr>
<tr>
<td><strong>Role</strong></td>
<td>Serving as a statistics specialist to Ken.</td>
</tr>
<tr>
<td><strong>Scenario operations</strong></td>
<td>Asking the expert for relevant information, running simulations or watching worked examples to verify the ideas.</td>
</tr>
<tr>
<td><strong>Resources</strong></td>
<td>Experts’ explanation about the relevant concepts and prerequisite knowledge, worked example, and simulations.</td>
</tr>
<tr>
<td><strong>Feedback</strong></td>
<td>Receiving positive or negative feedback after advising Ken.</td>
</tr>
</tbody>
</table>

There are three points in the development of *Statistics Specialist*. First is interviewing four experienced statistics instructors to gain understanding about issues
related to learning sampling distributions. The interview questions included: How important are sampling distributions? How difficult are sampling distributions? How do you teach sampling distributions? What parts of the instruction do the students find most difficult? And how do you know students have learned the concept of sampling distributions? Interview results were analyzed and provided the researcher valuable ideas and suggestions for design of the instructional program.

The second point is the preparation of learning materials. The researcher investigated four introductory statistics textbooks, which offered a template for organization and a description of concepts. These books are *Statistics for the behavioral and social sciences: A brief course*, by Aron, Aron, & Coups (2005); *Statistics for the behavioral sciences*, by Gravetter & Wallnau (2004); *Statistical methods for psychology*, by Howell (2002); and *The basic practice of statistics*, by Moore (2004). Part of resources about teaching sampling distributions came from research studies (e.g. Chance et al., 2004; delMas et al., 1999; Mills, 2000; Yu et al., 1995).

Chance et al. (2004) proposed that prerequisite knowledge to learning about sampling distributions should include the concepts of variability, a distribution, the normal distribution, and sampling. In addition, they listed various aspects of sampling distribution that students should understand (Appendix A), which were taken into consideration for developing *Statistics Specialist*. To comprehend what functions a simulation for sampling distributions should possess, the researcher reviewed similar computer simulations about sampling distribution online, such as Rice Virtual Lab in Statistics (http://onlinestatbook.com/stat_sim/sampling_dist/index.html), Statistics Online Computational Resource (http://www.socr.ucla.edu/htmls/SOCR_Experiments.html), and
Consortium for the Advancement of Undergraduate Statistics Education (http://serc.carleton.edu/sp/cause/datasim/examples/18169.html).

The final task was development of the instance of the GBS model, Statistics Specialist (Appendix J). Based on the suggestions from statistics experts and the review of relevant studies, the instruction was divided into six questions, ranging from fundamental to more complicated (Appendix B). Each question was embedded into the actor-narrated scenarios, which described the scenario’s problem (managing a shrimp farm). The actor, “Ken,” presented questions to the learner through short video clips. If the learners did not feel confident in giving Ken advice, they could either choose to “Run the simulation” or “Ask the expert” (Figure 1). In the “Ask the expert” section, users could choose questions to ask an expert and, once clicked, a video clip would play to deliver the instruction. In the “Run the simulation” area, learners could decide either to watch a worked example first or to try the simulation. When users acquired the necessary information to reply to the questions, they clicked the “Advise Ken” button to give “Ken” advice.

Statistics Specialist received two formative evaluations by two pilot studies. The first pilot study was conducted on October 11, 2008. Eight graduate level students were recruited from an instructional design class to examine the program. Participants had 30 minutes to evaluate the tutorial and they spent 11 minutes on average. Results indicated that participants tended to answer the questions by guessing while advising the client in the simulation. In addition, some program faults (e.g. button and graphic) were discovered. After revising the errors and adding a function that required a user to review essential resources in order to prevent irresponsible usage, the researchers invited two
graduate students (whose majors were Instructional Technology, and Educational Research and Evaluation) to individually test *Statistics Specialist*. Their suggestions included adding a function that allowed learners to selectively review certain sections of video clips, rather than having them replay the whole clip to catch up what they miss.

![Diagram of the learning path of Goal-Based Scenarios Computer Simulation](image)

*Figure 1.* The learning path of Goal-Based Scenarios Computer Simulation

The researcher resolved these problems and conducted the second pilot study in January 2009. Eight subjects from different majors were recruited to evaluate the program and each one received fifteen dollars to compensate his or her time. Similarly,
the participants had 30 minutes to evaluate the program. The results (Appendix C) revealed two major issues: insufficient time and unclear user guidance. Three participants indicated that they needed more time to finish the program. Two participants pointed out that they still needed more explanation of what options were required and what were not, although they had been notified at the beginning of the first question.

The researcher revised the design problems based on participant suggestions from these two pilots and then sent the introduction of *Statistics Specialist* to Roger Schank (personal communication, February 23, 2009), who proposed the GBSs model. His reply message indicated that *Statistics Specialist* fulfilled the fundamental ideas of the GBSs model except for the flaw of placing a time limitation on usage (Appendix D). Thus, the researcher removed the time limitation to faithfully follow the GBS model.

*Collect and Analyze Formative Data on the Instance.*

A critical advantage of conducting qualitative research is the opportunity to use various sources of evidence, such as documents, interviews, and observations (Yin, 2009). Using multiple resources not only allows researchers to investigate a comprehensive range of issues and facilitates discovery of converging perspectives of inquiry, such as the process of triangulation (Patton, 2002; Yin, 2009). Most researchers agree that analysis of data is one of the most difficult parts in qualitative research. The following sections introduce methodology for collecting data, procedure, and data analysis.

*Data Collection Method*

Two techniques were adopted to collect formative data: observations and interviews. Observations offer researchers the opportunity to understand how participants
experience and interact with the unexpected or expected in the program (Glesne, 2006). Participants’ answers, as Glesne (2006) indicated, can be accurately interpreted if interview questions developed through observations are connected to their known behavior. Schatzman and Strauss (1973) categorized approaches of conducting observations into six types: (1) watching from outside (the field researcher observes without being seen); (2) passive presence (the field researcher is present in the situation and avoids influencing the event); (3) limited interaction (the researcher intervenes in the subject’s ongoing event in order to make clarifications); (4) active control (the researcher actively controls the interactions in the event to collect desired data); (5) observer as participant: the researcher is a full participant in the event with a known identify as a researcher); (6) participation with hidden identity (the researcher is a full participant in the event with a hidden identify as a researcher). In the present study, the observation strategy was one of limited interaction.

Interviewing participants is the most direct and useful way to collect formative data. The purpose is to “allow us to enter into the other person’s perspective” (Patton, 2002, p. 341), which helps researchers identify strengths and weaknesses in the design instance. Interview questions, as Reigeluth and Frick (1999) suggested, should focus on discovering particular aspects of the implementation of the program, such as what elements work well, what does not work well, and ways to improve the weakness. In addition, if participants encounter problems while performing the task, the researcher should provide assistance before they proceed. This action may save future data from being corrupted by earlier flaws in the instance (Reigeluth & Frick, 1999).
Three methods are used when conducting interviews: think aloud, debrief, and focus group. The think aloud approach requires participants to verbalize their thoughts while they are performing a task (Ericsson & Simon, 1993). Ericsson and Simon (1993) found that some subjects may misunderstand the instruction and thus produce trivial information, such as providing irrelevant small talks and explaining or describing the process to the experimenter. To reduce these problems, Ericsson and Simon (1993) offered some guidelines: (1) an interviewer should make sure that the experimental situation is arranged not to allow social interaction; (2) an interviewer should offer instruction and practice problems that allow subjects to understand the procedure and the normative content of think-aloud verbalizations; (3) an interviewer should be seated behind the subject and clearly warn participants against explanation and verbal description.

Debriefing interviews occur after participants finish the study treatment. Data is collected on participant usage of the instance as a whole. Denscombe (2007) differentiated types of interview into structured, semi-structured, and unstructured. Structured interview imposes tight control over the questions and answers to gain direct points of interests from respondents, such a questionnaire conducted fact to face with an interviewee. In an unstructured interview, informants’ thoughts are the emphasis and they are offered maximum flexibility to provide information. Semi-structured interview embraces the strengths of both approaches; the interviewer can ensure that the critical inquiry is pursued and be free to word questions in order to acquire in-depth information (Patton, 2002). In this study, the researcher adopted semi-structured interview since it
allows to elicit the participants’ responses in a more open-ended and less structured approach.

A focus group interview refers to a group of five to ten people discussing specific issues (Krueger & Casey, 2009). Patton (2002) indicated the difference between one-on-one interview and focus group interview:

In a focus group, participants get to hear each other's responses and to make additional comments beyond their own original responses as they hear what other people have to say. However, participants need not agree with each other or reach any kind of consensus. Nor is it necessary for people to disagree. The object is to get high-quality data in a social context where people can consider their own views in the context of the views of others. (p. 386)

Thus, interactions among participants in a focus group present a more natural context that not only engages the participants in active comparisons of the opinions (Morgan, 1998) but also enhances the quality of the data (Krueger & Casey, 2009). In addition, researchers (Morgan, 1998; Patton, 2002) stated that focus group interview, to some extent, provides instant assessment of shared or diverse perspectives from interviewees.

Data Collection Procedure

In this study, participants came to the research room at an individually scheduled time. The researcher briefly introduced the study purpose and procedures. Participants completed a consent form, a demographic survey, and a pretest (Appendix E). Before the treatment, participants were informed that the researcher was testing an instructional approach and wanted them to critique it. They were told that their comments were very
important since the researcher would identify strengths, weaknesses, improvements of the instructional program on the basis of their responses. They were encouraged to be critical; any problem they encountered would be the weakness of the approach, rather than their deficiencies.

During the treatment, think aloud and observation methods were employed to gather data. Two computers were set up for observation. The first one was installed with *Statistics Specialist* and screen capture software (SnagIt 7) to video record the screen while participants used the program. In addition, an unobtrusive video camera was set up on the first computer, but connected to the second computer, to prevent overload on the first computer and allow observation of participant’s facial expressions without intruding. This helped the researcher to ask questions. For example, when the participant smiled during the treatment, the researcher would ask the participant to think aloud and verbalize his or her thoughts. When the participants encountered problems while using the program, the researcher directly provided assistance with the hope that subsequent data would be saved from corruption by earlier flaws in the instance (Reigeluth & Frick, 1999). All spoken comments were recorded to insure thoroughness of data collection. Also, the researcher noted his observations of the participant working with the simulation.

After treatment, participants took a posttest (Appendix F) and then the researcher asked additional interview questions. The debriefing questions were a different set of questions from those in the think aloud phase (Appendix G). Video clips that recorded their facial expressions and computer screen during the treatment were presented to help participants describe their experiences with *Statistics Specialist*. When the interview
ended, the researcher thanked the student for his or her participation. All the processes mentioned above took place during the first phase of data collection.

The second data collection session was a focus group interview that took place right after transcribing all the individual interviews. The researcher emailed the participants to schedule the most agreeable time. During this data collection session, participants first took a retention test (Appendix H) and then everyone, including the researcher, gave a short self-introduction to evoke conversation. A slideshow of *Statistics Specialist* was presented to remind participants of their experiences. Acting as moderator, the researcher asked questions, kept the conversation on track, and made sure that everyone had a chance to share. Video and audio recorders were used to record the whole interview. Some questions were asked to confirm the findings of the one-on-one interviews, which may enhance the external validity of the study. When the focus group interview completed, participants reviewed their own case reports of the previous individual interviews and the researcher thanked them for participation.

*Data Analysis*

Data analysis begins when data collection is ongoing (Glesne, 2006; Reigeluth & Frick, 1999). Data analysis is a dynamic process that a researcher has to ruminate, try out different ideas, and expand upon others before coming to any conclusion (Corbin & Strauss, 2008). Maxwell (2005) suggested two steps to analyze qualitative data. The first is reading the interview transcripts, observational notes, and relevant documents. The analyst should write notes and develop ideas about categories and relationships among the data while reading. The second step refers to the usage of analytic approaches, such as memos, categorizing strategies, and connecting strategies. Writing memos not only
helps the researchers capture their analytic thoughts, but also assists them in developing, stimulating, and reflecting insights. The most common use of categorizing strategies in qualitative research is coding (Maxwell, 2005). Coding allows the analyst to analytically break the data apart and rearrange them into categories (Strauss, 1987), which helps to develop a more specific focus or more relevant question (Glesne, 2006). Coding can also facilitate the researcher’s organizing the data into broader themes, patterns, or issues (Maxwell, 2005). Connecting strategies are ways to identify relationships among the data to be coherent as a whole, instead of breaking them apart.

In this study, data was gathered through interview, observation, field notes, pretest, posttest, and a user tracking mechanism which recorded the time while the user clicked each button in Statistics Specialist. The researcher used qualitative analysis software, MAXQDA (http://www.maxqda.com/), to sort, connect, code, and display the resulting data in order to identify strengths, weakness, and possible revision of the GBS model.

Revise the Instance

Reigeluth and Frick (1999) stated that researchers can revise the instance as soon as they feel confident in their value and use the revised instance for the remaining data collection. They can acquire participant opinions about use of both versions of the design (before and after the revision). Given the complicated nature of revising a computer simulation, it is not practical to make changes to the program during implementation. Results of two pilot studies were used to identify and revise errors in the program itself before the study began. The instance described in this study was not revised during its implementation.
Repeat the Data Collection and Revision Cycle

Reigeluth and Frick (1999) recommended that within the boundaries of the theory, the cycle of data collection, analysis, and revision should be repeated as much as possible. By doing so, researchers can confirm earlier findings and also identify more information for revisions, which may enhance external validity (generalizability) (Reigeluth & Frick, 1999). This study represents the first round of data collection, analysis, and revision.

Offer Tentative Revisions for the Theory

At this stage, researchers can use their findings to propose an improved instructional design model. But these suggestions merely present recommendations for revisions to the model. It still takes future studies that completely replicate the model and revisions to validate the modifications (Reigeluth & Frick, 1999). In the present study, the researcher offers a tentative revised version of the GBSs model at the conclusion of this dissertation.

Methodological Issues

Probably from historical bias or personal preference, qualitative research has been criticized for the lack of rigor (Davies & Dodd, 2002). In order to deal with this concern, Reigeluth and Frick (1999) addressed three methodological issues: construct validity, sound data collection and analysis procedures, and attention to generalizability to the theory.

Construct Validity

The focus of construct validity, according to Yin (2009), is “identifying correct operational measures for the concepts being studied” (p. 40). That is to say, “the concepts
of interest in formative research are the methods offered by the design theory, any situations that influence the use of those methods, and the indicators of strengths and weaknesses” (Reigeluth & Frick, 1999, p. 647). The concepts of interest in this study include: the approaches provided by the GBS model and the indicators of strengths and weaknesses of the GBS model.

Reigeluth and Frick (1999) indicated two threats to construct validity: excluding elements that are in the model, and including elements that do not belong to the model. The best way to avoid these issues is to have experts in the theory assure the construct validity (Reigeluth & Frick, 1999). In this study, every effort was made to follow the model as much as resources allowed during development of Statistics Specialist. In addition, the GBS model expert, Roger Schank, agreed on the methods in Statistics Specialist, except placing a time limitation on usage. This flaw was corrected before the initiation of the present study.

Yin (2009) recommended three tactics to increase construct validity: using multiple sources of evidence, establishing a chain of evidence, and having the case study report reviewed by key informants (member checking). In the present study, formative data was collected through multiple sources, such as observations during participants use, a series of one-on-one interviews, and a focus group interview. All the data gathered was organized and inspected to identify and establish links among resources. Finally, the draft of the interview transcript was reviewed and confirmed by the participants. The researcher may further interview with subjects to correct errors or misconceptions.

Another factor that may influence construct validity is the validity of instructional content. Or in other words, were learning materials able to transmit valid information to
the participants? To offer sound instruction, the researcher prepared the content based on
the opinions of four statistics experts’ and a review of relevant studies and textbooks.
During the development process, the researcher constantly discussed the content with
these experts and they checked the soundness of the materials when finished.

*Sound Data Collection and Analysis Procedures*

Sound data collection and analysis procedures are influenced by two major factors
(Reigeluth & Frick, 1999): the completeness of the data and the accuracy of the data. The
former can be enhanced by several techniques, such as advance preparation of
participants, an emergent data-collection process, gradually decreasing obtrusivity (Dick
et al., 2005), iteration until saturation, and identification of strengths as well as
weaknesses. The latter can be improved using triangulation, a chain of evidence, and
member checking, which are already discussed as techniques to enhance construct
validity. Among them, triangulation plays a critical role to compare and cross-check the
consistency of the data. Patton (2002) proposed some suggestions to conduct
triangulation of data sources, including:

1. Comparing observations with interviews;
2. Comparing what people say in public with what they say in private;
3. Checking the consistency of what people say about the same thing over time;
4. Comparing the perspectives of people from different points of view. For
   example, during an evaluation, a researcher could triangulate staff views,
   client views, funder views, and views expressed by people outside the
   program; and
5. Checking interviews against program documents and other written evidence that can corroborate what interview respondents report. (p. 559)

*Advance Preparation of Participants*

Although there was no relationship between the researcher and recruited participants, students were informed that the researcher was testing a new approach and wanted them to critique it. Any problem they encountered would be the weakness of the approach, rather than their deficiencies.

*An Emergent Data-collection Process*

Since the researcher had little idea of the weaknesses and areas of improvements in the GBSs model, open-ended probes and flexible data collection were used in this study, instead of adhering to the researcher’s lines of questions.

*Gradually Decreasing Obtrusivity*

The researcher used an obtrusive approach, think aloud, to gather data while a student is using the simulation. But the obtrusivity will fade in later rounds to confirm the earlier findings with later subjects.

*Iteration Until Saturation*

Each participant represents a different iteration of the data collection, which allows the researcher to confirm prior findings. The researcher will continue to collect data until the prior findings are confirmed.

*Identification of Strengths and Weaknesses*

This is the focus that is driving data collection and analysis in this study. What works well and what does not will be explicitly investigated.
Attention to Generalizability to the Theory

The generalizability to the theory can be done by recognizing situationality and replicating the study. Situationality, according to Reigeluth and Frick, refers to ways that activities, events, and methods may vary for different situations. In the present study, the researcher explored situationalities when the results turned out differently in each round of interview. In addition, the researcher examined different learners (students who took statistics courses vs. students who never took any statistics courses) to see if results differed. “When situationalities are incorporated into the theory, the theory becomes useful for a broader range of situations” (Reigeluth & Frick, 1999, p. 649).

The present study is the first evaluation of the GBSs model. Suggestions proposed cannot directly modify the GBSs model until future studies confirm the findings of this study using Statistics Specialist implemented in a similar format. Additional studies may also need to investigate the application of the GBSs model in a different context and conduct an evaluation. Only when sufficient replications are completed does the proposed improvements to the design model gain sufficient evidence to warrant changes in the model (Reigeluth & Frick, 1999).

Participants

Participants in this study were ten graduate students with a little or no prior statistical knowledge. They were recruited either from specific classes (e.g. Introduction of Research Methods and Social Structure & Change in Education) or through the poster advertisements. Table 2 presents the interviewees demographic information. As seen in the table, students came from five different majors and the females were younger than the males. The study consisted of two sections, individual interview and focus group
interview. Ten students, five females and five males began Part 1 of the study, then two male students dropped out because the first one was unable to come on the date that all the students agreed on. The researcher recruited a second one

Table 2.

Demographic Information of Participants

<table>
<thead>
<tr>
<th>Participant</th>
<th>Major</th>
<th>Gender</th>
<th>Age</th>
<th>International Student</th>
<th>Date</th>
</tr>
</thead>
<tbody>
<tr>
<td>Participant 1</td>
<td>College Student Personnel</td>
<td>Female</td>
<td>18-24</td>
<td>No</td>
<td>4/17/09</td>
</tr>
<tr>
<td>Participant 2</td>
<td>College Student Personnel</td>
<td>Female</td>
<td>18-24</td>
<td>No</td>
<td>4/17/09</td>
</tr>
<tr>
<td>Participant 3</td>
<td>College Student Personnel</td>
<td>Female</td>
<td>25-30</td>
<td>No</td>
<td>4/19/09</td>
</tr>
<tr>
<td>Participant 4</td>
<td>Athletic Training</td>
<td>Female</td>
<td>18-24</td>
<td>No</td>
<td>4/22/09</td>
</tr>
<tr>
<td>Participant 5</td>
<td>Athletic Training</td>
<td>Female</td>
<td>18-24</td>
<td>No</td>
<td>4/24/09</td>
</tr>
<tr>
<td>Participant 6</td>
<td>Environmental Studies</td>
<td>Male</td>
<td>25-30</td>
<td>No</td>
<td>4/29/09</td>
</tr>
<tr>
<td>Participant 7</td>
<td>Counselor Education</td>
<td>Male</td>
<td>Over 40</td>
<td>No</td>
<td>4/24/09</td>
</tr>
<tr>
<td>Participant 8</td>
<td>African Studies</td>
<td>Male</td>
<td>Over 40</td>
<td>Yes</td>
<td>4/23/09</td>
</tr>
<tr>
<td>Participant 9</td>
<td>Counselor Education</td>
<td>Male</td>
<td>31-40</td>
<td>No</td>
<td>5/08/09</td>
</tr>
<tr>
<td>Participant 10</td>
<td>Athletic Training</td>
<td>Male</td>
<td>25-30</td>
<td>No</td>
<td>5/04/09</td>
</tr>
</tbody>
</table>
to replace him, but he still forgot to attend Part 2 study. Thus, For Part 2 of the study, two males dropped out, so that eight students, five female and three males, completed the study. The participants who dropped out Part 2 study received an email from the researcher to have them review their own case reports of the previous individual interviews.

The Pilot

A pilot project was conducted on April 2, 2009 to gain insights on the data collection procedure and to examine the stability and quality of data recorded from software and equipment. The interviewee was a graduate student majoring in Instructional Technology who took three statistics courses in more than one year. During the pilot, the participant stopped verbalizing her thoughts while watching the instructional movies. The researcher did not prompt her to express her thoughts so as to not disturb the learning activity. The appropriate timing for prompts occurred when the participant was selecting resources (such as choosing Ask the expert or Run the simulation), after watching the video clips, advising the client, or running the simulation.

The researcher found that sometimes the participant would ask the researcher questions during the thinking aloud process. For example, while choosing the accurate advice for the client, the interviewee verbalizing her thoughts, “I know the first option is more accurate than the third one, but I am wondering why Option 3 is wrong. What do you think?” The researcher did not give her explanation but encouraged her to focus on the instruction. The interviewee pointed out some confusing design issues while expressing her thinking. Instead of asking for her reasons and suggestions, the researcher wrote down the occurring time on the field note and allowed uninterrupted continuance
of the thinking aloud process to prevent impedance of the learning process. When the participant finished the treatment, the researcher presented the video clip to the designated time and asked the participant’s explanation.

Observations and observations techniques were examined in this pilot. In the beginning of the treatment, the participant expressed discomfort with the camera. The researcher again explained the purpose of capturing her facial expressions and reassured her that the data would be protected. The researcher watched the participant’s facial expression from another computer and prompted her with questions when facial changes were observed. For example, when the participant received the first positive feedback from Ken, the researcher asked her to verbalize her thinking and she replied that Ken’s feedback was a good encouragement. Overall, the audio and visual data recorded had good quality and the data collection devices were running stable. The placement of the devices was appropriate.

Summary

In this chapter, the researcher introduced the methodology of formative research that was used to improve the GBS model. The detailed methods of implementing formative research and improving the rigor and validity of this study were provided. Chapters 4 and 5 present the results of the data analysis and findings of this study.
CHAPTER 4: FINDINGS

Introduction

This chapter presents the results of the study as related to the research questions: “What components in the Goal-Based Scenarios (GBS) model worked and which did not work?,” and “What improvements might be made?” To evaluate the GBS model, the researcher first developed a tutorial, or instance, based on the model and named it Statistics Specialist. It was designed to teach the concept of sampling distributions. Ten participants were recruited to critique the instance and improve the GBS model. Data were gathered through three interview techniques: think aloud, debrief, and focus group. The findings were organized into the framework of a GBS model, which includes goal, mission, cover story, role, scenario operation, resources, and feedback. Each section first describes how the component was implemented in this study, and then displays the evaluation of what did and did not work well. The last section presents the results of the pretest, posttest, and retention test to demonstrate how students’ understanding improved by using Statistics Specialist.

Learning Goal

What Happened?

A learning goal is what the designer would like the students to learn. It falls under two categories: 1) process knowledge that practice skills to achieve the goal, and 2) content knowledge that seeks and figures out the targeted information to attain the goal. Since GBS models focus on improved outcomes, the learning goal in Statistics Specialist was to teach students a statistical concept, sampling distributions, which was the primary learning goal. Sampling distributions are fundamental concepts that are necessary for
testing statistical hypotheses. Many studies (delMas et al., 1999; Chance et al., 2004) indicate that students often report difficulty and confusion while learning sampling distributions, especially the abstract process of randomly and infinitely taking samples from the population.

*What Went Well and What Didn’t?*

A GBS becomes feasible when it arouses students’ interest and when both designer and learner share the same goals. In this study, when participants were asked about their attitudes toward learning statistics, most expressed fear, anxiety, and a noted lack of enthusiasm. As Participant 3 said, “I’m not exactly thrilled about learning it because I’m a little nervous when it comes to numbers. It’s been a long time since I’ve had a math class.” A similar response was offered by Participant 6: “When I did the graduate level research, I tended to stay away from statistics and mathematic or formulas and stuff like that as much as I could. Because statistics is not something everybody easily understands.” Participant 8 explained, “I felt statistics didn’t mean anything to me. I thought it was very difficult to understand and very complicated.”

Although most participants were not interested in learning statistics, some participants still pointed out the importance of studying statistics. Participant 1 said, “I think it [statistics] is very important to know, but I also wish that I knew more than I do. And I feel like I should know more.” Similarly, Participant 2 said:

I know that I need to learn statistics but I don’t realize what they’re really saying. I know I should learn it and I have no problem trying to figure it out if I have to. It is a little scary.
Participant 3 thought that understanding statistics would help her reading research papers, saying that, “It would be very good for me to be able to learn it [statistics] and understand it. Later when I am reading research, I can understand when they are talking about.”

Despite participants’ stated disinterest in learning statistics, it did not seem to overly influence their usage of *Statistics Specialist*. When asked about their experiences working on the tutorial, student feedback was positive. For example, Participant 1 said, “I definitely like it. I think it is better than sitting down and reading a book to try to learn it. I felt a lot of confidence after I went through that and it made me feel smarter.” She further stated:

I never take any stats class[es]. It triggered things that I heard before, like standard deviation, but I couldn’t remember what it was. And then I remember[ed] what it was when I went through it [*Statistics Specialist*]. But I think it will be helpful because especially now I am going to a research class. I feel I have a better grasp of what I am going to do. I think it was interesting. It was fun. It didn’t bore me.

Participant 6 indicated that he learned key concepts by using *Statistics Specialist*, saying:

It was a good experience. I learned something. When I did the pretest, I didn’t understand any of them. In the posttest, I understood. Although I didn’t get them all right, at least I got the fundamental understanding of what sample means are and behavior of sampling distributions.
Mission

What Happened?

Once the goal of a GBS design is determined, the second step is creating a motivational mission that promotes a desire for the user to pursue. A mission is an activity that a learner practices through the application of targeted skills or knowledge in the simulated scenario. Once learners complete the mission, they can achieve the goal, which means that they have acquired the concept, which in this case, was sampling distributions. In *Statistics Specialist*, the mission was designed for the learner to assist the client Ken measure shrimp weight.

What Went Well and What Didn’t?

When participants started *Statistics Specialist*, they were instructed to play a role within the mission as a statistics specialist to assist their client Ken. The researcher found that not every participant was aware of the mission (Table 3). Among those who perceived the existence of the mission, they valued the importance of the mission. As Participant 4 indicated:

> It is important. It is a business he is running and he wants to find the best way to go about doing so. So, your job being the analyst [statistics specialist] is to help him do that. That is what he comes to you for.

Some participants expressed a sense of pressure from the mission and a perceived burden to acquire the necessary information in order to advise their client. As Participant 7 pointed out:

> I felt some pressure. Pressure that, gosh this guy’s coming to me for help and I’m really not an expert. So, I need to go ask somebody and learn how to understand it
so I can give him some good advice. So, I felt some pressure. Hey, I need to learn this stuff to help Ken.

Some participants agreed that the mission was motivational (Participants 4, 7, 8, & 9) and similar to real-life experiences (Participants 4 & 5).

Table 3.

Participant Distribution of Mission and Role Awareness

<table>
<thead>
<tr>
<th></th>
<th>Mission</th>
<th>Role</th>
</tr>
</thead>
<tbody>
<tr>
<td>Participant 1</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>Participant 2</td>
<td>X</td>
<td></td>
</tr>
<tr>
<td>Participant 3</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>Participant 4</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Participant 5</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Participant 6</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>Participant 7</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Participant 8</td>
<td></td>
<td>X</td>
</tr>
<tr>
<td>Participant 9</td>
<td></td>
<td>X</td>
</tr>
<tr>
<td>Participant 10</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

X, indicates that participants were not aware of the element.

Some participants indicated that they were unaware of their mission and focused more on learning materials than performing the mission. Participant 1 stated, “I didn’t think much about the mission. I was using the tutorial to answer the questions and to find
out his [Ken’s] ultimate answer.” She further said “I didn’t know it [the mission] so much matter. I thought it could be just any situation to try to answer the question. I was more interested in the information. So, I was learning rather than helping.” Participant 3 expressed a similar opinion, “It makes sense. I think it is good choice. It is something visual you can see. It is simple, easy to understand…. I don’t think it is about the mission. I think it is about the concept.”

Cover Story

What Happened?

The third component is GBS cover story. The cover story is the scenario that creates the needs for the mission. Like the mission, it should be motivational and realistic so learners willingly accept the role and are inspired to further pursue the goal. The cover story must provide sufficient opportunities for users to practice the targeted skills or seek out target knowledge.

The background storyline in Statistics Specialist describes a man named “Ken” who has a new job working on a shrimp farm. Ken’s boss is currently trying a food supplement to accelerate shrimp growth. In order to assess supplement effectiveness, she wants Ken to constantly monitor shrimp growth, and more specifically, measure the average weight of the shrimp in the pond.

The pond contains more than 100,000 shrimp, so Ken understands that it is impossible to measure every shrimp in the pond. He randomly draws samples of 30 shrimp and computes the mean for each sample. Due to difference among these sample means, he seeks an explanation from the user who plays the role of a statistics specialist. This is the first question that Ken asks the users, and six questions were developed to
teach three essential characteristics of sampling distributions: center, spread, and shape. These questions are presented from simple to complicated.

*What Went Well and What Didn’t?*

The intent of the cover story was to create the need for the mission to be accomplished. After using the tutorial Participant 1 said, “I didn’t know much about shrimp, but I think adding the story in to have the whole mission work toward, instead of just having random questions that don’t match any common goal. I like how they are all related.” Some participants pointed out that the cover story could make them learn in a realistic situation. For instance, “I thought it made it more applicable. But instead of just turning on the simulation and learning the information, it [the cover story] made it [the tutorial] seem like how you could actually use the information in real-life” (Participant 2). Furthermore, some students indicated that the cover story provided a good storyline to make the application of the targeted knowledge concrete. Participant 3 said:

> It is a method to get information about it. So, the story is okay. I think it is easy, something visual or something easy to see. If you pick a different number like age or weight or something, it may not be quite as concrete as measuring shrimp.

Participant 10 agreed and said, “It helped me imagine the concept because shrimp is one situation where you can’t actually pull out the whole population when you have 10,000 or 20,000 shrimp. So it makes sense to use something like shrimp.”

Some participants pointed out that they generated inquiries when Ken prompted them with questions (Participants 4, 7, 8, 9 & 10). As indicated by Participant 4, these questions included, “What can I do to help him? What is the best way to go about figuring this out? What is exactly he asking me for?” She further stated:
This is a real-life situation, so you want to figure out how you can help him. It is motivational to think, what can I do to solve these problems? And I am not somebody who likes leaving problems unsolved. So, it makes me want to figure it out and find a way to go about doing it.

Participant 8 expressed the same thought that “I think that with simulations like this one [I] can better appreciate statistics than [with] traditional methods. When you get into statistical issues, it may become very abstract. You seem to compare and to relate practical issues with statistical methods.”

Although most participants agreed that the cover story provided a good problem-solving situational context for them to apply their newly acquired target information, a few participants indicated disinterest in the shrimp farming story. As Participant 6 said, “It is a good problem applying problem, but I don’t know. I just don’t think whether people care about shrimp…. I am just wondering about how interested people would be learning about shrimp, shrimp sizes, and the distribution of shrimp.

Participant 3 and 7 felt similarly. Participant 7 said, “It wasn’t real exciting for me, but the same concept can be applied to other things too. You could do candy bars or you know some things other people are interested in.” In addition, Participant 9 thought that the cover story helped him least, because it did not help him solve the problems. It did help him understand the background information given by the story, however.
Role

What Happened?

The role is the character that the learner assumes within the cover story. It should be the one who practices the necessary skills or applies the targeted information in the scenario. Not only can the role provide learners with a specific set of expectations as to how, when, and why they are to demonstrate their knowledge and skills. The role also provides motivation for them to be engaged in the application. In Statistics Specialist, the students’ role was to serve as a statistics specialist to the client, Ken, and help him seek appropriate explanation to his questions.

What Went Well and What Didn’t?

When asked if assuming a role provided sufficient motivation for learning, some participants gave an affirmative answer. Participant 2 said: “I will definitely say yes. It [playing the role as a statistics specialist] is more stressful to get the right answer.” Another participant elaborated, “It’s really important that when someone is relying on you, you give them the best and most accurate advice that you can. So, I tried to pay as best attention as I could to the explanations that were given” (Participant 9). The same theme was further developed by Participant 7:

I think it [the role] was really good to have that term specialist going in this [tutorial]. I didn’t feel like a specialist, but they’re using that term to put the pressure on you and want you to do well. You don’t want to let Ken down, you know, being called a specialist here. He’s looking at you for answers, so it did put pressure on you to focus and try to do your best job.
The scenario also helped some participants gain confidence. Participant 4 explained:

I think it [the role] gave you a little bit of automatic confidence as well. Because instead of being called a student or something like that, you automatically have that title as a specialist, so you kind of feel like not power but like you already have some knowledge. So, it can help you better.

Participant 5 felt that the title of the role was encouraging, saying “I think it’s motivating. It helped like you were held to a standard, so you kind of wanted to meet that standard”.

Despite instructions at the beginning of the tutorial to assume the role of the statistics specialist, most participants shared that they did not take on the role. Possible explanation for this behavior can be divided into two types. Type one is simply that they forgot the role. Participant 1 said, “Well, I forgot that I was a statistics specialist.” Participant 3 offered further explanation stating, “I kind of forgot about it to be honest. After it was all done and you [the researcher] asked me about it I had totally forgotten about that slide. I guess I wasn’t into it and [the role] didn’t mean anything to me”.

Participant 3 pointed out that the scenario role-playing as a statistics specialist did not influence her learning since she did not consider herself a statistics specialist. Participant 6 expressed a similar opinion:

I was kind of indifferent. It wasn’t until the end when you asked the question in the first part where I hadn’t even really considered about my part, you know, my role in the study [Statistics Specialist]. I hadn’t really considered. So, I guess I’m all certain indifferent.
The second possible explanation for students not assuming the role of statistics specialist was the perceived pressure of being a statistics specialist. This is thought to be the result of self-induced pressure by students who know they are not true statistics specialists and not qualified to provide accurate advice, thereby refusing to accept the role in the tutorial. One participant explained, “Well, I felt like I didn’t want to fail the client or give him wrong information. I felt pressured and I wanted to do a good service” (Participant 7). Another said, “It’s kind of overwhelming at first just because I don’t consider myself a specialist in much. I thought it was kind of funny that they called me that when I barely had a clue about what I was doing” (Participant 9). Similarly, Participant 8 stated, “I didn’t get my role as statistics specialist very seriously because I knew I was guiding something. But I was very far from being a specialist.”

Participant 8 even reported discomfort with the ‘statistics specialist’ title. He explained: “I was not very conscious about my role as a specialist. In fact, in the end of the exercise, when he replied to me as an expert [in a letter to show appreciation the user], I thought I was just being teased”.

Due to the pressure of assuming the specialist role, some participants ignored the role and paid more attention to acquisition of the targeted information. “I thought if I gave him the wrong answer…. I felt it was kind of more pressure to know the right answer. So, every now and then I guess my role is more about learning the information” (Participant 2).

Some participants thought that the feeling of pressure might improve with more time spent on the tutorial. Participant 10 said:
[Playing the role as a statistics specialist takes] a lot of responsibility. At first, it was scary because I didn’t know that much about it. But as it went on, you know, it gave me a sense of empowerment because I knew and I had the title. I knew what was going on. I knew what to answer and how to answer Ken’s question. It made it easier.”

Scenario operations

*What Happened?*

Scenario operations consist of the activities that students undertake to accomplish the mission and achieve the goal. They must be closely related to the goal and mission and must include decision points with consequences. In *Statistics Specialist*, the scenario operations included asking the expert for relevant information and running the simulation to explore or validate related ideas. The simulation offered a platform for an important step in statistical learning, which is the comprehension of randomness and approximation (Chance, Ben-Zvi, Garfield, & Medina, 2007). A simulation gives the user a chance to manipulate parameters in order to make abstract concept concrete through direct observation. Participants were allowed to select their favorite resources. Decision points were designed to require users to give the client accurate advice. Whenever participants felt confident and ready to give Ken advice, they clicked the “Advising Ken” button. A positive feedback, such as “Good job, that sounds reasonable to me,” was displayed if the learner picked up the right answer. Otherwise, a negative feedback, like “Mmmm, that is not right. You need to ask the expert or run the simulation again,” was provided to inform them that they did not clearly understand the information. The feedback was presented in the form of video clip and narrated by a female guide named Ashley.
What Went Well and What Didn’t?

Participant 4 pointed out that the scenario operations were helpful. She indicated that with the goal, mission, and Ken’s questions in mind, she felt empowered by the encouragement offered by scenario operations, which said “It’s up to you to go on and to find and use the resources. You have to come up with the answer.” Some participants shared that the scenario operations provided a realistic experience (Participants 4, 8, & 10). As Participant 4 stated:

It is good and it is real-life as well. Because when somebody comes to you with a question, for the most part you are going to have time to do your research, look into it, and get background information on it before you get back to the person and give them the answers. When somebody asks you something, you are not just going to give them something off the top of your head. It is like doing your research.

Participant 1 expressed a similar response and said, “Actually by going through movies [the video clips in Asking the expert section], doing the sample [run the simulation], and answering the questions [advise Ken] helped reiterate all the information and made me feel like I was actually trying to find my answers in the right way.” Further, Participant 10 pointed out that using the scenario operations was as realistic as working in his job as a counselor, saying that:

If we receive a client that we don’t know how to serve that client or we need more advice, we go to the expert. We may actually look up past clients with the same situation and advice there. So, I think it definitely helps out to ask somebody else to get their advice. Definitely helps out.
In addition, Participant 6 thought that the scenario operations could satisfy learners with different learning styles. He indicated:

I think that is good for people with different learning styles because everybody doesn’t have the same learning styles. When you have options for people, they can, you know, have more than one way to lead them to the answer. I think that is good.

Participant 4 liked the control over the resources in the tutorial, saying:

I thought it was very helpful. I liked how it gave you options. You didn’t have to sit through the whole tutorial if you didn’t feel like you needed it. If you felt you could go right ahead and gave him your advice, then that is helpful as opposed to have you sit through the whole tutorial where it could cause frustration for some people, I think.

Not every participant completed all the resources before advising Ken. But the researcher found that the participants went through all the resources while answering Ken’s first question. After the first question, they would choose resources according to their preference. If they did not learn from the resource in the beginning, they tended to not use it or use it after. This was explained by Participant 2:

I went to the first button and ran the simulation and then ask the expert. I understood better from “Ask the expert” because I didn’t exactly know what I was looking at for running the simulation like I thought I did…. And then I decided I was only using one.

Similarly, Participant 9 pointed out that, “When I got the question, I asked the expert first. If I couldn’t answer the question from there, then I might run the simulation and
then after that went back to Ken.” None of participants tried out the answers by guessing and they all went back to review the resources when they gave Ken wrong advice. Conversely, and probably due to lack of motivation, participants in the pilot study tended to guess the answers without reviewing the learning materials.

Resources

What Happened?

Resources provided well-organized and accessible information for the users to accomplish the mission. In Statistics Specialist, resources included “Ask the expert” and “Run the simulation”. When the learners selected “Ask the expert”, they were provided with several questions to which the expert could reply through video clips. These questions were relevant to the critical concept in Ken’s questions and provided prerequisite knowledge about the idea. The participants could pause, replay, or play a specific part of the clips while watching the movies. In “Run the simulation”, the learners could choose either “show me how to do it” that presented a worked example or “let me try it” that allowed the users to manipulate the parameters and observe the results through the simulation. There was no time limitation on using the resources and participants were allowed to switch between these two resources as well as resources offered in previous questions.

What Went Well and What Didn’t?

All participants pointed out that resources were the most helpful of the seven components. As Participant 1 pointed out:
I went to this [Statistics Specialist] not knowing anything, so I definitely needed the resources to be able to get to the next step. If I didn’t have the resources, I would just be guessing and trying to figure out what the definition was.

Students agreed on the sufficiency of resources and the advantage of repetitively using the resources until they understood the concept. They thought that the information was clear and that the sequence and amount of content were well-distributed. Some students liked the way that simulations visualized the concepts, while other favored watching the video clips to acquire the information. When asked about their preference for the resources, except Participant 4 who favored “Run the simulation”, most students preferred “Ask the expert” for three reasons:

1) Lecture approach.

The participants indicated that they were accustomed to the lecturing approach in the video clips because it was similar with the way used in the traditional lecture, having an instructor stand in the front and teach the concept (Participants 1, 2, 3, 5, 6, 7, 8, & 9). As Participant 2 indicated, “I felt it was just like the way we were used to being taught, like having someone teach it rather than me reading it.” In addition, there were two types of clips in “Ask the expert.” One mainly presented graphics of the simulation together with the instructor’s voice-over. The other showed the instructor and his voice-over. When investigated their preference of two types of video clips in “Ask the expert,” some students (Participants 2, 3, 5, & 6) still liked the latter because of seeing an actual person in the clips.

2) Duplicate Content
All video clips in “Ask the expert” already included the critical concepts in “Run Simulation” and some of them presented graphics captured from the simulations. As a result, the students thought that watching these video clips in “Ask the expert” was sufficient for them to understand the concept, making it unnecessary to run the simulation (Participants 3 & 5).

3) Individual Preference

Participant 5 preferred the “Ask the expert” since she thought that she could get the definition of everything from asking the expert, which satisfied her learning style by acquiring definitions first. Participant 7 tended to use “Ask the expert” because the graphs in “Run simulation” were intimidating. Participant 1 preferred “Ask the expert” because she liked the concept that was broken down more detailed than that in “Run the simulation”. In addition, she had a preference that someone walked her through the instruction, rather than trying to figure it out by herself.

Participants who preferred “Run the simulation” liked how the simulations visualized the concept and how they interacted with the simulations to construct the meaning (Participant 3 & 4). Participant 4 said:

I guessed I tended to go to run the simulation after the first couple of problems because it was easier for me to see what the actual curve might look like and what information I could pull from that as opposed to just listening to the expert…. Just because you could actually see what you were doing. Either they could do it for you and you watch or you read the instruction and figure out yourself.

Participant 3 further pointed out that:
I thought it was helpful for me to run the simulation and I loved being able to plug my own numbers in. Because I felt like when you’re telling me what numbers to plug in, you know, it’s getting the response that you wanted. And I didn’t really understand how it worked. So, instead of putting 10 in, I was going to put 50 in and to see how it works and whether it came out with the same result or a very similar result. That was good for me.

Despite individual preference, some participants agreed on the value of using both resources. Particularly, when they did not understand their preferred one, they would seek the other resource. As Participant 7 said, “I tended to go to the expert first because the graphs [Run the simulation] were intimidating. And then I found out I needed to look at the graphs, too. So it was good to have both.” Participant 8 thought that these two resources had close relation, saying “I think there is a close connection between these two. In order to appreciate one, you have to relate to the other one.” Participant 1 shared a similar comment, “Going to the “Ask the expert” allows me to get the background information and I can go the simulation with a better idea of what I am going to be doing, rather than just jumping right in.” When asked about their feeling after taking the posttest, Participants 3 and 6 indicated that in order to get the questions right, it was necessary to do “Ask the expert” and “Run simulation.” Table 4 shows the participants’ suggestions about the resources.

Feedback

*What Happened?*

Feedback is given through the consequence of choices that learners make in the scenario. In *Statistics Specialist*, feedback was either presented as encouragement
whenever the users offered Ken accurate advice, or given as remediation feedback to inform them that they misunderstood the information. According to Schank et al. (1999), identifying the gap in understanding may motivate students to pay more attention to the targeted content. Another type of feedback was given when the users answered all of Ken’s questions. It is a letter narrated by Ken to show his gratitude for the specialist’s essential information.

Table 4.

<table>
<thead>
<tr>
<th>Participant</th>
<th>Suggestion</th>
</tr>
</thead>
<tbody>
<tr>
<td>Participant 3</td>
<td>Provide a glossary to quickly review the critical terms.</td>
</tr>
<tr>
<td>Participant 4</td>
<td>Summarize the main points in each video clip in “Ask the expert”.</td>
</tr>
<tr>
<td>Participant 6</td>
<td>Modify wording or slow down the narration in “Show me how to do it” of Question 2.</td>
</tr>
<tr>
<td>Participant 7</td>
<td>Make the cover story more appealing and replace the text instruction in “Let me try it” with narration.</td>
</tr>
<tr>
<td>Participant 8</td>
<td>Provide at least two different explanations for each concept taught in the tutorial.</td>
</tr>
<tr>
<td>Participant 10</td>
<td>Offer more extra questions in “Ask the expert” and make this section more interactive as well.</td>
</tr>
</tbody>
</table>
What Went Well and What Didn’t?

Participants thought that the feedback functioned well and gave them emotional support. Participant 1 explained, “[Positive feedback] made me feel good when I knew the answer. It was nice to have somebody say the feedback, rather than just seeing, you got it right. I think it added a more personable and realistic factor in that.” In addition, Participant 8 revealed that feedback enhanced his confidence, saying “It [Positive feedback] was giving me greater confidence; particularly, the feedback from every operation [question]”. Similarly, Participant 9 said that “I like that [the feedback] because it’s always nice to get rewarded for when you do good things. So, I think that was a good part of it”. Participant 7 agreed and pointed out, “It [feedback] is kind of a cheerleader there, cheering you on when you do well. And it’s good feedback when you messed up here, saying that you might want to look at this section again. I mean that’s very positive.”

When participants gave wrong advice to Ken, they received feedback telling them that they answered incorrectly and needed to review the content again. This often elicited an emotional response and motivated them to do better (Participants 1, 2, 5, 7, 8 & 9). As Participant 1 said:

I was bummed because I thought I had a grasp on it. It [The feedback] motivated me to go back and pay more attention to make sure that I knew what the definition of those words were and make sure I was reading the question thoroughly. So, it definitely did motivate me because I want to get the right answer.

Participant 5 expressed that she was motivated, saying, “That was bad. I get competitive, so I don’t like being wrong. So, I get upset.”
However, not every participant thought that feedback was necessary (Participants 3 & 6). As Participant 3 said, “I thought it was kind of cheesy. I almost laughed the first time because it didn’t really do anything for me. But it’s good to have because some people might need it. I’m motivated enough.”
When asked about how feedback could be improved, Participant 1 suggested that it would be helpful to show a hint when they gave Ken wrong advice. Participant 4 pointed out that offering reviewing information while giving positive feedback would be useful as well.

Results of Understanding

The ultimate goal of a GBS model is the facilitation of learning. A good approach to examine the achievement of this goal is through the evaluation of participants’ understanding, which consists of a pretest, posttest, and retention test. Before the interview about their usage of Statistics Specialist, the researcher asked participants some questions to probe their understanding. The content below is a transcript of the interview with Participant 3.

Researcher: If I draw a sample of size 30 from a population, calculate the mean, and repeat this process many many times, could you tell me how the sample size influences the distribution of sampling means?

Participant 3: If the sample size you are drawing is larger, it [the mean of sampling distributions] will be a better representation of what the actual population is. So, the means [of sampling distributions and the population] should be close to each other. But if the sample size you
are drawing is small, it could be scattered anywhere. So you are not sure where it is.

Researcher: What does the shape of sampling distributions look like?

Participant 3: The mean [of sampling distributions] will be equal to the population mean. The shape will be normal, even though the population is not normal.

An investigation of data revealed that some participants still had difficulty putting the learned concepts together. For example, Participant 4 could not predict the shape of sampling distributions when given a small sample size and an abnormal population.

Researcher: If I draw a sample of size 2 from a U-shape population, calculate the mean, and repeat this process many times, the distribution of sample means is gradually forming. What does the shape look like?

Participant 4: It will be more spread-out since the sample size is small. It could be a normal distribution.

The researchers found that some participants were unable to offer proper explanation and tended to use rote memorization while using Statistics Specialist.

Researcher: If I draw a sample of size 5 from a normal population, calculate the mean, and repeat this process many times, could you tell me where the mean of sampling distributions locate?

Participant 7: So the mean would be in the middle there.

Researcher: Is the mean the same with the population mean?

Participant 7: I don’t know.
Researcher: Ok, if the sample size is 35, what do you think the mean of sampling distributions will be?

Participant 7: I think it’s in the center.

Researcher: Could you tell me why?

Participant 7: …. 

Researcher: Ok, what will the shape of distribution look like?

Participant 7: If more than 30, it would be normal is what I’m retaining.

Table 5 displays how each participant spent time on *Statistics Specialist* and on each scenario operation. Table 6 provides the descriptive information of the data. As seen in Table 6, the average time spent on the tutorial was about 46 minutes. Students spent more time on “Ask the expert” and less time watching worked examples (Show me how to do it).
Table 5.

*Time Spent on Resources*

<table>
<thead>
<tr>
<th>Participants</th>
<th>Ask the expert</th>
<th>Show me how to do it</th>
<th>Let me try it</th>
<th>Tutorial</th>
</tr>
</thead>
<tbody>
<tr>
<td>Participant 1</td>
<td>22 min 08 sec</td>
<td>3 min 18 sec</td>
<td>16 min 36 sec</td>
<td>53 min 34 sec</td>
</tr>
<tr>
<td>Participant 2</td>
<td>27 min 18 sec</td>
<td>3 min 54 sec</td>
<td>4 min 46 sec</td>
<td>50 min 04 sec</td>
</tr>
<tr>
<td>Participant 3</td>
<td>23 min 41 sec</td>
<td>3 min 20 sec</td>
<td>5 min 55 sec</td>
<td>42 min 09 sec</td>
</tr>
<tr>
<td>Participant 4</td>
<td>5 min 57 sec</td>
<td>3 min 56 sec</td>
<td>17 min 20 sec</td>
<td>35 min 16 sec</td>
</tr>
<tr>
<td>Participant 5</td>
<td>19 min 24 sec</td>
<td>10 min 13 sec</td>
<td>10 min 17 sec</td>
<td>48 min 23 sec</td>
</tr>
<tr>
<td>Participant 6</td>
<td>17 min 49 sec</td>
<td>10 min 52 sec</td>
<td>1 min 27 sec</td>
<td>38 min 11 sec</td>
</tr>
<tr>
<td>Participant 7</td>
<td>12 min 30 sec</td>
<td>2 min 32 sec</td>
<td>14 min 55 sec</td>
<td>45 min 43 sec</td>
</tr>
<tr>
<td>Participant 8</td>
<td>29 min 06 sec</td>
<td>14 min 14 sec</td>
<td>5 min 44 sec</td>
<td>69 min 08 sec</td>
</tr>
<tr>
<td>Participant 9</td>
<td>18 min 05 sec</td>
<td>4 min 16 sec</td>
<td>1 min 49 sec</td>
<td>31 min 25 sec</td>
</tr>
<tr>
<td>Participant 10</td>
<td>24 min 51 sec</td>
<td>5 min 54 sec</td>
<td>10 min 09 sec</td>
<td>51 min 19 sec</td>
</tr>
</tbody>
</table>

Table 6.

*Descriptive Information for Time Spent on Resources*

<table>
<thead>
<tr>
<th></th>
<th>Minimum</th>
<th>Maximum</th>
<th>Mean</th>
<th>SD</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ask the expert</td>
<td>5 min 57 sec</td>
<td>29 min 06 sec</td>
<td>18 min 15 sec</td>
<td>7 min 00 sec</td>
</tr>
<tr>
<td>Show me how to do it</td>
<td>2 min 32 sec</td>
<td>14 min 14 sec</td>
<td>5 min 50 sec</td>
<td>4 min 03 sec</td>
</tr>
<tr>
<td>Let me try it</td>
<td>1 min 27 sec</td>
<td>17 min 20 sec</td>
<td>8 min 05 sec</td>
<td>5 min 53 sec</td>
</tr>
<tr>
<td>Tutorial</td>
<td>31 min 25 sec</td>
<td>69 min 08 sec</td>
<td>42 min 17 sec</td>
<td>10 min 46 sec</td>
</tr>
</tbody>
</table>
All participants completed a pretest and posttest. Both tests included eight multiple-choice questions. Students were required to indicate their confidence level with their choice when finished each question. Table 7 and Table 8 show the mean and standard deviation of the pretest, posttest, and confidence level for each answer. As shown on both tables, females performed better on the pretest and posttest than males, but males were more confident in their answers than female on both tests.

Table 7.

*Results of Pretest*

<table>
<thead>
<tr>
<th></th>
<th>Mean</th>
<th>SD</th>
<th>Mean of Confidence Level</th>
<th>SD of Confidence Level</th>
</tr>
</thead>
<tbody>
<tr>
<td>Female</td>
<td>3.8 (47.5 %)</td>
<td>1.1</td>
<td>31.8</td>
<td>15</td>
</tr>
<tr>
<td>Male</td>
<td>2.4 (30 %)</td>
<td>0.9</td>
<td>48.8</td>
<td>10.2</td>
</tr>
</tbody>
</table>

Table 8.

*Results of Posttest*

<table>
<thead>
<tr>
<th></th>
<th>Mean</th>
<th>SD</th>
<th>Mean of Confidence Level</th>
<th>SD of Confidence Level</th>
</tr>
</thead>
<tbody>
<tr>
<td>Female</td>
<td>6.8 (85 %)</td>
<td>1.3</td>
<td>68.5</td>
<td>18.7</td>
</tr>
<tr>
<td>Male</td>
<td>5 (62.5 %)</td>
<td>1.6</td>
<td>79.8</td>
<td>5.7</td>
</tr>
</tbody>
</table>
Eight students participated in the second part of the study, which included a retention test before the focus group interview. Differing from the items in the pretest and posttest, the retention test included four questions that required students to provide justification for their answers (Appendix H). The researcher found that most participants had difficulty offering accurate explanations, although they could answer the questions correctly. Table 9 presents their explanations to Question 1, which asked, “What is the distribution of sampling means?” As shown on the table, most participants were unable to answer this question accurately, except Participant 2.

Table 10 displays participants’ explanations to Question 2, which asked students to predict the mean of sampling distributions that were randomly drawn from a unimodal population with a sample size of 10. With the exception of Participants 3 and 7, participants could correctly indicate the mean of sample distributions, 60.5, but not everyone could offer reasonable justification. As seen on Table 10, only Participant 8’s response roughly grasps the key concept of the question.
<table>
<thead>
<tr>
<th>Participant</th>
<th>Explanation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Participant 1</td>
<td>How the mean of the samples is distributed across the graph for each sample size.</td>
</tr>
<tr>
<td>Participant 2</td>
<td>The distribution of sampling means is the mean that represents a population. By taking what are known as samples or sampler groups within the population that are randomly selected to represent the entire population.</td>
</tr>
<tr>
<td>Participant 3</td>
<td>This is a confusing question (what does distribution of sampling mean?). Distribution of sampling has to do with the way you collect your sample for your test.</td>
</tr>
<tr>
<td>Participant 4</td>
<td>Where the means lie on a graph and the pattern that they make. Depending on the pattern that can show if there is a normal or skewed distribution.</td>
</tr>
<tr>
<td>Participant 5</td>
<td>Distribution of the sampling means would be all the samples taken from the population or sample and then would be put on a graph, so you would be able to see.</td>
</tr>
<tr>
<td>Participant 6</td>
<td>The distribution of sampling means is normal.</td>
</tr>
<tr>
<td>Participant 7</td>
<td>No explanation offered.</td>
</tr>
<tr>
<td>Participant 8</td>
<td>The distribution of sampling means is the range at which data are spread within a graph.</td>
</tr>
</tbody>
</table>
Question 3 tested their understanding of the relationship between sampling distributions and the population in terms of the variability, shape, and sample size.
Although five participants got the question right, only three of them (Participants 2, 3, & 4) provided reasonable justifications (shown in Table 11).

Table 11.

<table>
<thead>
<tr>
<th>Participant</th>
<th>Explanation</th>
<th>Results</th>
</tr>
</thead>
<tbody>
<tr>
<td>Participant 1</td>
<td>A. The samples were done in a large number and many times.</td>
<td>Incorrect</td>
</tr>
<tr>
<td>Participant 2</td>
<td>B. Because if the sample size is above 30, it is a normal distribution.</td>
<td>Correct</td>
</tr>
<tr>
<td>Participant 3</td>
<td>B. With a larger sample size (larger than 35) the distribution of sample means narrows down to the mean of the total samples. This is tough for me to explain.</td>
<td>Correct</td>
</tr>
<tr>
<td>Participant 4</td>
<td>B. Because the larger the sample is the more compact the graph will be.</td>
<td>Correct</td>
</tr>
<tr>
<td>Participant 5</td>
<td>C. The larger sample size will get an increase of the samples.</td>
<td>Incorrect</td>
</tr>
<tr>
<td></td>
<td>You have a better chance of getting more diffuse distribution with the sample size of 50 rather than 2.</td>
<td></td>
</tr>
<tr>
<td>Participant 6</td>
<td>A. the distribution of sample means becomes more normal as sample size increases.</td>
<td>Incorrect</td>
</tr>
<tr>
<td>Participant 7</td>
<td>B. B is the sample.</td>
<td>Correct</td>
</tr>
<tr>
<td>Participant 8</td>
<td>B. The sample size of 50 is more likely to be even.</td>
<td>Correct</td>
</tr>
</tbody>
</table>
Question 4 focused on how sample size from different shapes of the population influences the shape of sampling distributions. Table 12 presents their responses. Among the five participants who answered this question right, only Participant 2 and 6 offered sound explanations.

Table 12.

<table>
<thead>
<tr>
<th>Participant</th>
<th>Explanation</th>
<th>Results</th>
</tr>
</thead>
<tbody>
<tr>
<td>Participant 1</td>
<td>D. It can be either depending on the sample.</td>
<td>Correct</td>
</tr>
<tr>
<td>Participant 2</td>
<td>D. When the sample size is above 30, it is a normal distribution which could have come from any of the histograms.</td>
<td>Correct</td>
</tr>
<tr>
<td>Participant 3</td>
<td>D. It depends on if the means are evenly distributed or not. I don’t really remember much more.</td>
<td>Correct</td>
</tr>
<tr>
<td>Participant 4</td>
<td>C. Because A and B do not show a mean of 60. C shows a normal distribution with mean 60.</td>
<td>Incorrect</td>
</tr>
<tr>
<td>Participant 5</td>
<td>D. All the means are 60 and the sized sample of 50 would be similar for the distribution of each.</td>
<td>Correct</td>
</tr>
<tr>
<td>Participant 6</td>
<td>D. The distribution of sample means always approaches normal distribution as sample size increases, no matter what the distribution of the population is.</td>
<td>Correct</td>
</tr>
<tr>
<td>Participant 7</td>
<td>C. C is normal.</td>
<td>Incorrect</td>
</tr>
<tr>
<td>Participant 8</td>
<td>C. No explanation offered.</td>
<td>Incorrect</td>
</tr>
</tbody>
</table>
Summary

This chapter provided results from implementing the GBS model in the form of the statistical tutorial, *Statistics Specialist*. Ten masters-level students participated in Part 1 of the study and only eight students completed Part 2. Results showed positive results that highlighted strengths of the GBS model, including:

1. Engaging students in the learning topic even though they indicated that they were not interested in learning statistics.
2. Learning goal, mission, role, and feedback were interdependent and enhanced motivation.
3. Students favored the cover story that prompted questions and allowed them to choose their favorite resources for searching relevant information.

This chapter also identified some issues about using *Statistics Specialist* that need to be improved. They are summarized below:

1. The assignment of the mission and the role in the beginning of *Statistics Specialist* imposed stress on some participants because they did not understand the targeted information yet.
2. Feedback was superficial since it merely told the users whether they got it right or wrong, rather than directing them to relevant information.
3. Most participants were unable to offer reasonable explanations to questions asked in the retention test, although their posttest scores improved from the pretest.

A discussion of the results in terms of their implications for GBS is presented in the next chapter.
CHAPTER 5: DISCUSSION AND CONCLUSION

This chapter begins with a discussion of data analysis results in terms of strengths and weaknesses of Goal Based Scenario (GBS) components, and offers tentative recommendations within the context of existing research literature. In addition, this chapter evaluates the study as a whole and indicates study limitations. The last section discusses the implications of the findings and suggests areas for future research.

The intent of this study was to identify strengths and weaknesses of the GBS (1993/1994) instructional design model so that the model might be improved. This study used the qualitative research methodology, formative research, to investigate and answer the following questions which were the focus of the study:

1. What are strengths and weaknesses of the goal-based scenario model?
2. What improvements might be made?

The researcher designed an instance, Statistics Specialist, based on the GBS model to teach the concept of sampling distributions. Participants were recruited to critique the instance as a way to further evaluate the GBS model. The researcher used think-aloud, debrief, and focus group interviewing techniques to collect data.

Discussion and recommendations of GBS strengths and weaknesses are organized by component. Strengths are strategies or tactics that the participants considered helpful and worked well for them, whereas weaknesses are things that they viewed as not helpful or ineffective. Weaknesses that identified possible improvements for the GBS model are provided as well.
Strengths of GBS

**Learning Goals**

The statement of learning goals showed the direction of the learning activity and was identified as a strength. Four participants indicated that they wanted to learn statistics in order to conduct research or read research papers. They seemed to be motivated in *Statistic Specialist* when the learning goal resonated with their need to learn, although they expressed no specific interest in the learning goal.

This finding supports Knowles, Holton, and Swanson’s (2005) adult learning theory proposing that “adults need to know why they need to learn something before undertaking to learn it” (p. 64). They are more motivated to learn if they perceive that learning will help them perform tasks or solve problems that they may encounter in life situations (Knowles et al., 2005). In fact, goals influence performance through four mechanisms (Locke & Latham, 2002): 1) directing attention and effort to goal-relevant activities, 2) having an energizing function, 3) affecting persistence, and 4) indirectly affecting action by leading to arousal, exploration, and strategies. From this perspective, the statement of learning goals in *Statistics Specialist* seemed to be effective.

**Mission**

The learners identified the mission as a source of motivation that gave them a sense of investment necessitating that they acquire requisite knowledge before advising the client (Participant 4, 5, 7, & 8). This finding is in agreement with the perspective of Schank et al. (1993/1994) proposing that the main responsibility in the mission is to encourage a sense of investment in the GBS. When learners have ownership over some aspect of their GBS experience, they are engaged and driven to complete the mission.
Cover Story

Participants described the cover story as a helpful element. They felt that the cover story provided a storyline that connected the mission and other components as a whole. Furthermore, offering a story that included prompts for user questions provided varied opportunities to learn target skills and knowledge. Some participants indicated that the shrimp farming story made the conceptual application of sampling distributions more concrete and applicable, demonstrating how they could actually apply the information in real-life.

In addition, participants identified the questions asked by the client as a motivational way to generate inquiries and drive them to seek relevant information. This finding supports Jonassen’s (1999) claim that in any constructivist learning environment, questions that students attempt to solve will drive learning. Meaningful learning begins when one generates inquiries and tries to make sense out of learning materials (Mayer, 2002). Further, learners are more motivated to learn when the problems prompted are situated in natural context (Brown, 1989). In perspective of mathemagenic effects, the cover story prompting questions functioned as a way to provide mathemagenic information that offered adjunct questions to effectively stimulate and facilitate the attainment of instructional objectives (Rothkopf, 1970).

Role

Participants stated that assuming the role in Statistics Specialist was motivational. Some expressed the view that helping the client was a good source of motivation that prompted them to give the client the most accurate answers they could. Some felt that the title would somehow give them confidence to solve the problems. Still others indicated
that assuming the role of the specialist seemed to set up a standard for them to meet. These findings support theoretical claims that role-play allows the development of role expectation and promotes engagement in learning activities (Nikendei et al., 2005; Van Ments, 1999). As indicated by Resnick (1998), participation of role-playing activities enables students to ‘dive in’ mathematical and scientific phenomena and to develop relations with the knowledge underlying the phenomena.

**Scenario Operations**

Participants identified scenario operations as strengths. These operations consisted of “Ask the expert”, “Run simulation”, and “Advising Ken”. Participants mentioned that it was practical and helpful to run scenario operations together with questions prompted in the cover story, and that it seemed a natural approach to deal with daily problems. This finding supports Schank and Cleary’s (1995) natural learning theory in which learning is achieved through the process of adopting a goal, generating a question, and developing an answer.

Additionally, results showed that learner control was helpful in terms of allowing them to select their preferred resources, keep their own pace, and repeat sections where they felt unclear. Some participants reported that learner control satisfied different learning styles. When asked about their own learning styles (independent learners tend to explore information first by themselves while learning something new, whereas dependent learners are inclined to receive instruction first), those who claimed dependent learning styles tended to prefer using “Ask the expert”; the only one considering herself an independent learner liked to use “Run simulation.” The dependent learners revealed that they selected “Ask the expert” since the video clips could walk them through the
concept, instead of trying to figure out by themselves. The independent learner preferred manipulating the numbers with step-by-step instruction and seeing the simulated graphics in “Run simulation.” Some dependent learners liked to ask the expert first and then run the simulation later to validate the concepts.

The ability to control learning not only enabled the instructional system to be adapted to the users’ preferences and cognitive levels (Merrill, 1980), but also facilitated students’ knowledge construction and development of self-regulatory skills (Scheiter & Gerjets, 2007). Students tend to evaluate consequences associated with self-directed learning and they learn how to learn, if learning control is provided (Barab, Bowdish, Young, & Owen, 1996).

**Resources**

Nine out of ten participants identified resources as the most helpful component to promote their understanding of sampling distributions. They agreed that the resources were concise, clear, and accomplishable with reasonable effort and felt that the content was broken down in a simple-to-complex sequence, which played an important role in the quality of the simulation (Reigeluth & Schwartz, 1989). Some participants felt that with support from the resources, they gained more confidence in answering the client’s questions. This finding supports the theoretical claim that scaffolding can promote cognitive self-efficacy (Van Eck, 2007). When the learners’ perceived competence is increasingly supported with timely feedback, their perception of their own learning abilities will also improve and be motivational (Bandura, 1997).
Feedback

After providing accurate advice to the client and receiving feedback, some participants gained a certain degree of encouragement and confidence (Participant 1, 2, 5, 8, 9, & 10). Even the feedback they received while offering wrong advice still influenced participants’ emotions and motivated them to review the relevant information (Participant 1, 5, & 7). This finding supports Bandura’s social cognitive theory (1986), claiming that goals enhance motivation through self-reactive influences. As he further explained:

When individuals commit themselves to explicit goals, perceived negative discrepancies between what they do and what they seek to achieve create self-dissatisfactions that serve as incentives for enhanced effort. The motivational effects do not derive from the goals themselves but rather from the fact that people respond evaluatively to their own behavior. Goals specify the conditional requirements for positive self-evaluation. The more self-dissatisfied people are with substandard performances, the more they heighten their efforts (p. 469, as cited in Bandura & Cervone, 1983, 1986; Locke, Cartledge, & Knerr, 1970).

The combination of learning goal with timely feedback seems to produce an emotional engagement that enhances the learning activity.

Summary

Responses from some participants pointed out that all GBS components were important and each component depended on each other to promote engagement and enhance learning. The mission derived from the learning goal assigned a task for learners to pursue. The cover story provided a realistic context and opportunities to practice targeted skills and knowledge. The element of role-play engaged students in learning
activities. Scenario operations, resources, and feedback scaffolded learning. Combined, these components embody the process of equilibration. That is, cognitive disequilibrium was created and support was provided to achieve equilibrium.

A good way to investigate the effectiveness of a GBS model as a whole is through an evaluation of participants’ understanding. Posttest results showed that participants’ performance and confidence levels increased. In addition, most participants could offer explanation with only slight errors when given an open-ended question during a debriefing interview. These results provided additional support that the GBS model worked well.

**Improvements in GBS: Changes**

Based on data analysis results and reported weaknesses and suggestions, each component in GBS, except learning goals, needed to be changed in order to improve the model.

**Mission**

The data revealed that participants experienced stress while performing the mission in the beginning of *Statistics Specialist* since they did not know much about the concepts yet. This finding is related to Bandura’s (1986) claim about self-efficacy. He described that people tend to spend time thinking about how they perform while starting a task. Those with a strong sense of efficacy attend to task challenges and generate a competent attitude toward incoming scenarios, whereas those with a weak sense of efficacy focus on personal deficiency and generate negative attitudes toward any challenge. Encountering stress, according to Piaget, can be a good source of motivation
that drives students to restructure their knowledge (Wadsworth, 2004). The technique is to maintain the stress level within learners’ zone of proximal development (ZPD).

According to Reeve (2005), a possible way to increase self-efficacy is to empower people through self-efficacy training that employs a mastery modeling program in which an expert shows learners how to deal with the problems that lowers their self-efficacy. In computer-based instruction, this is done by offering a worked example that models task performance behaviors and provides support.

Role

The role was the least helpful component in GBS according to three participants during the debriefing interview and five during the focus group interview. Many of whom reported that they did not take on the role while using Statistics Specialist. Researchers (Aldrich, 2005; Rollnick, Kinnersley, & Butler, 2002; Swink, 1993) have shown that role-playing activity may form barriers for some students, especially adults, who for various reasons may be unfamiliar or unwilling to buy into the experience. In this study, the reasons for not assuming the role could be attributed to two factors: overlook and refusal. The former described participants who paid more attention to the acquisition of knowledge and little to assuming the role. The latter described participants who felt that they knew nothing about the concept yet, so did not consider themselves as a specialist.

Both issues are probably due to inadequate introduction. In Statistics Specialist, the researcher merely displayed a short description (Figure 2) and then played the video clip of the cover story, hoping that the participants could ‘dive in’. Apparently, this was an insufficient stimulation for participants’ involvement in the role play. Researchers
(Vallius, Kujanpää, & Manninen, 2006) have pointed out that playing a role could be a learning process and through these experiences, role-playing ability might be improved.

A possible solution for these two problems is to provide a video clip modeling the role-playing behavior as well as instruction that introduces the ideas for which the role play was designed to support. This may allow users to be acquainted with the operation of the tutorial in the beginning of usage. As Rogers (2007) indicated, the debriefing process is part of learning since students can understand the objectives and rules from observation. When learners are familiar with the situation in role play, they can understand the role and the potential responses or behaviors (Lowenstein, 2007).

From the perspective of motivation, encouraging the pursuit of learning goals instead of performance goals in the debriefing clip may be a possible way to reduce students’ anxiety of assuming the role. Learning goals, as indicated by Schunk (2008),
emphasize learners’ attention on learning processes and strategies that help them construct knowledge and improve their skills, whereas performance goals focus on the completion of the tasks. Students of learning goal orientation tend to make positive attributions for success and sustain their self-efficacy for learning (Bandura, 1993).

**Scenario Operations**

No participants pointed out any weakness in the scenario operations. Although while examining the participants’ test results, the researcher found that the use of multiple-choice questions in “Advising Ken” seemed less challenging for the participants and promoted rote learning (Ausubel, 2000).

Scenario operations, according to Schank et al. (1999), should be constructed with decision points that either signify successful completion of the mission or failure. Learners acquire targeted knowledge when the mission is achieved. An investigation of test results showed that on average, participants correctly answered 90% of questions in Statistics Specialist. During the interview, some participants said that those questions were uncomplicated. However, although participants performed well when answering questions within the GBS program and during the posttest, more than half were unable to offer an accurate explanation when answering an open-ended question during the interview. Similarly, they could correctly answer the questions in the multiple-choice retention examination, but failed to offer a sound reason for the options they chose when asked to write a justification. While examining their responses, the researcher also found that most participants could merely identify one or two characteristics of sampling distributions, such as understanding how sampling size influences the spread or shape of sampling distributions. They still had difficulty integrating critical concepts and
explaining the process. According to Chance et al.’s (2004) developmental model on sampling distributions, most participants belonged in the transitional reasoning stage.

The usage of a multiple-choice examination in Statistics Specialist may have promoted shadow learning. For example, when Participant 6 was asked why he did not run the simulation, he replied that “I retained that information well, so I didn’t see the need to try it on my own.” Similarly, when Participant 7 was asked to predict the shape of the sampling distribution for a given sample size and population, the researcher found that he could only recite the rule, stating “More than thirty it would be normal is what I’m retaining” but he still couldn’t explain the reason. The finding supports Scouller’ study (1998) in which students were inclined to employ surface learning strategies and motives in the multiple-choice examination context and to perceive this type of examinations as assessing factual information (lower levels of cognitive processing).

Students who used superficial learning approaches tended not to perform well on the retention test. According to Ausubel (2000), rote learning may not result in the acquisition of any meaning and its retention may not last long. A good example of this was seen with Participant 3. She was the only one who correctly answered all the questions in “Advise Ken” and the posttest. She provided a sound explanation to the open-ended question in the debriefing interview, but was unable to offer any explanation in the retention test. Although other variables influence retention, such as learning time, poor retention can be partly attributed to the GBS design of multiple-choice examination that promoted shadow learning.

Possible remediation included adopting small group usage and promoting reflection with open-ended questions. There is no specific prescription for assessment in
GBS, though most GBS studies employed multiple-choice examinations. Given the lack of definitive assessment protocols, three ways are suggested to facilitate students’ engagement and create students’ disequilibrium while using Statistics Specialist.

1. Have two students work in a group and encourage discussion. This idea is based on social constructivism that encourages learners to not only collaboratively construct knowledge but also to support each other (Jonassen, 1999).

2. Design a field below each question and require learners to write an explanation. The purpose here is not to evaluate the correction of answer, which is still a challenge for instructional designers to do through computer-based instruction, but to allow them opportunities to reflect on what they have acquired. It is possible that students may provide unrelated or inaccurate answers, but at least they are prompted to retrieve information, organize it, and interact with activities.

**Resources**

While the data did identify resources as the most helpful element, the results also suggested changes that consisted of reducing information overload and supporting doing. First, participants expressed confusion about the terms they learned in the tutorial; some were newly acquired and some they learned previously, but had forgotten. Collectively, the terms imposed a large information burden. Offering a glossary or hot linked terms in the application to remediate this problem may be an appropriate approach since it not only allows learners to alleviate ignorance about a concept, but also quickly review the concepts already taught (Murray, Piemonte, Khan, Shen, & Condit, 2000).

Another information overload issue was demonstrated by one participant who felt distracted while reading text instruction and performing the simulation in “Run
A possible solution would be to replace text instruction with narration and allowing users to control each step they read. This suggestion is related to a modality effect that students understand narrated explanations and pictures more effectively than on-screen text explanations (Mayer & Moreno, 2003). Another participant recommended offering a summary in each video clip in “Ask the expert” section. Summary information was included for some clips, but not in the first two of the client’s questions, which were two that the participant watched. She further indicated that a summary would be helpful to reduce information overload. This finding supports the theoretical claim that providing a summary encourages learners to focus on relevant information (Mayer, 1999). GBS, when implemented with computer simulations, should follow Mayer’s recommendation.

Second, resources need to support doing. Participants in Statistics Specialist were provided with control over resources, which fulfills the constructivist ideal of individual knowledge construction (Duffy & Jonassen, 1992; Jonassen, 1999). The problem with this approach is the variability of how learners utilize specific resources, which may not be the way designers would like them to be used.

Despite the simulation’s capability to provide an excellent visualization of the abstract process of generating sampling distributions, most students still refused to use it. According to an investigation of the time spent on Statistics Specialist, participants tended to use “Ask the expert” more than “Run simulation.” The researcher designed the simulation because of evidence from numerous studies supporting the usage of simulations in statistics learning (Chance et al., 2007; Chance & Rossman, 2006; delMas et al., 1999; Garfield & Ben-Zvi, 2007; Mills, 2004). However, the participants’ preference to “Ask the expert” seemed to work against the intent of a GBS. This can be
seen in the study of Schank et al. (1993/1994), which compared GBS with anchored instruction:

GBSs differ from anchored instruction in an important way that relates to our use of the phrase goal-based. The role of the student in anchored instruction activities includes observing some events (e.g., a video), verifying the accuracy of the information, looking for clues in the materials, and applying those clues to solving a problem faced by some character (Cognition and Technology Group at Vanderbilt, 1990). In a GBS, on the other hand, the student becomes an active participant in the scenario. The student's motivation within a GBS is to move toward completing some task on his or her own behalf. We do not mean to imply that anchored instruction prevents students from assuming a participatory role within the simulation. Rather, we wish to highlight how central the student's role is in a GBS.

This description implies that an active participant in the GBS tutorial should engage in the learning activities provided in the instance, rather than merely passively watching video clips and applying the targeted information to solve the problem. In agreement with Schank et al., Schwartz and Bransford (1998) further indicated that, “Hearing a lecture may seem like a passive activity because novices often do not have sufficient background knowledge to approach the text generatively” (p. 510).

To help the learners be generative when listening to the lecture, doing is one of the essential factors (Schwartz & Bransford, 1998). Schank (2002) stressed, “In any case, listening should come after difficulties in doing rather than in preparation for doing” (p. 229). The usage of simulation in teaching sampling distribution underlies the idea of
learning by doing (Hagtvedt et al., 2007). It visualizes the forming of concepts by allowing students to manipulate parameters (e.g., change the shape of a population, choose different sample sizes, and compare different populations) and simulate sampling distributions by randomly drawing large numbers of samples.

A possible approach to remediate this problem is redesigning other GBS components, such as the cover story, role, or scenario operations, that intentionally limit usage in order to support doing-related activities that the subject experts consider critical. Within Statistics Specialist, when a user selects “Ask the expert”, the expert could direct explanation to the usage of simulation and provide narration with step by step guidance.

**Feedback**

The results showed that some participants identified feedback as the least helpful element in GBS, since it merely told the consequence of actions without offering any informative information. The feedback in GBS can be given in any of three ways (Schank et al., 1999): through the consequence of actions, through online coaching with just-in-time support, and through video clips in which domain experts tell stories of similar experiences. Because of limited resources, feedback was provided by displaying consequences of actions in Statistics Specialist. The purpose of showing learners the consequence of their choices, especially negative consequences, was to give them opportunities to reflect their experiences and construct their own knowledge. Schank et al. (1999) explained that:

> Once you experience an expectation failure, explanations become important. They form the lesson that you learn from the expectation failure. When something does not happen the way you planned, you need to figure out why that is. The reason
will help you to abstract a lesson that you can apply to your expectations in the future. (p. 171)

The participants in this study seemed to prefer direct guidance telling them what went wrong and what information to look at, which may hinder opportunities to reflect on their experiences and interact with learning materials. Further, it may merely promote the memorization of key facts rather than reviewing the material to gain a sound comprehension. Conversely, it is possible that brief feedback telling them that they are wrong may frustrate learners. As recommended by Participant 1, a possible solution for this dilemma is to display a hint describing which concepts to focus on. Another participant pointed out that positive feedback would be more helpful if it could recapitulate the critical concept in the question. This suggestion supports the theoretical claim that students can concentrate on the relevant information if summary is provided (Mayer, 1999).

Conclusions

The purpose of this study was to evaluate the GBS instructional design model by examining its designed instance, Statistics Specialist. Questions guiding this research included: What are its strengths and weaknesses?, and What needs to be improved? Formative research methodology was employed to gather and analyze data from ten graduate students. The data were based on the participants’ one-time usage of Statistics Specialist with an average usage time of 42 minutes.

The strengths of a GBS included: 1) learning goals that enabled learners to see their learning needs, 2) a sense of investment due to a mission that engaged students in the learning activity, 3) a cover story that provided a context and problems to enhance
students’ engagement in the program, 4) a role that increased users’ motivation through a title that the role inherits and through the aid of the client, 5) scenario operations that satisfied learning control and different learning styles, 6) Resources, indicated as the most helpful element, that promoted understanding, and 7) feedback that gave learners confidence and the perception of negative discrepancy that triggered further learning.

A GBS might become a better instructional design model if the following improvements are made: 1) provide a worked example or instruction that demonstrates the behaviors of using the program and seeking supports in order to increase the user’s sense of self-efficacy while pursuing the mission or assuming the role, 2) employ small group usage and open-ended question approaches to promote learners’ engagement and interaction in scenario operations, 3) carefully integrate other components in GBS to support hands-on activity, 4) display a hint with negative feedback and recapitulate the concept in positive feedback.

The ultimate goal of a GBS is to provide guidelines designing a simulated context that helps students take advantage of the software features and promote conceptual understanding. Although some weaknesses were identified in this study, these tentative recommendations were limited to the subject domain. More studies need to be conducted in order to confirm the findings of this study and provide valid evidence to achieve the ultimate goal of a GBS.

Recommendation for Practitioners

This study presents the implementation of the GBS model to teach a statistics concept and should provide insights for practitioners to develop and design a computer simulation using the GBS model. One strong recommendation emerging from this study
is to spend more time identifying methods that support and engage learners in doing-related activities. Learning by doing, the core tenet of a GBS, works “because it strikes at the heart of basic memory processes that humans rely on” (Schank, 2002, p. 5). However, evidence both from the literature and this study showed that it is challenging to involve adult learners in simulations.

Although role playing in a GBS allows users to learn by participating, making mistakes, and challenging themselves, practitioners should provide appropriate guidance to facilitate students’ engagement. Evaluation plays an important role in a GBS and should be provided with some way to allow learners to generate cognitive dissonance so that they can retrieve, organize, and construct knowledge. Finally, since a GBS instructional system begins with teaching a small concept that learners have difficulty conceptualizing, it is essential to select a reusable concept for software design. By doing this, future tutorials may be linked to previous ones, which saves time and effort.

Recommendations for Future Research

1. Additional research should investigate how students’ learning styles influence their usage of an instructional instance. What kinds of learning styles tend to engage in the tutorial? And how can a GBS be modified to satisfy learners with different learning styles?

2. Pretests should include items that assess students’ prerequisite knowledge to learn whether this knowledge influences their learning activity. In this study, although fundamental information about sampling distributions was offered in the program, some participants still expressed confusion and overload among these concepts, which may affect their usage.
3. Future studies should examine different usage of a GBS instance, such as working in pairs. The intent is to observe whether the strengths of social constructivism can be gained through collaborative simulation performance.

4. Since participants in this study were graduate level students, additional case studies should design and test an instructional instance using GBS with similar content and difficulty, but with undergraduate students in order to see whether similar results are obtained.

5. Future studies should recruit younger participants. The participants were more inclined to learn by listening (Ask the expert) than by doing (Run the simulation). A possible explanation is that these students had been schooled in a passive way of receiving information. They did not want to use the simulation, which requires more cognitive effort.

6. Additional case studies should design and test an instructional instance using suggestions proposed in this study with different content and learners in order to see whether the results of this study can be applied to other situations.

Limitations of the Study

1. Conclusions drawn from this study are tentative and will not become valid until they have been thoroughly replicated and validated by future studies.

2. Conclusions drawn from this study cannot be generalized to other content implementations beyond statistics education.

3. Participants’ prior knowledge may have played an important role in statistical learning, but was not investigated in this study. Although Statistics Specialist provided the fundamental concepts that were suggested by a previous study (Chance
et al., 2004), participants seemed to encounter information overload under the circumstance of limited usage and time.

4. Conclusions drawn from this study may be affected by age differences. On average, male participants were older than female participants and two male students were over 40. Although their performance in the posttest improved from the pretest, both participants had difficulty offering explanation to an open-ended question in debriefing interview and in the retention test. All participants expressed a strong sense of low efficacy in learning statistics.

5. The time between the posttest and the retention test was not unified among participants, since participants were randomly and individually recruited to receive the treatment.

The use of a single focus group slightly weakened data collection since about half of the participants were less active respondents. Despite the researcher’s effort to encourage discussion, some students offered little opinion.
REFERENCES


environments based on the blending of microworlds, simulations, and games.

Hill/Open University Press.


Ryan, A. (2000). Peer groups as a context for the socialization of adolescents motivation,
engagement, and achievement in school. *Educational Psychologist, 35*(2), 101-
111.

games and simulations: A systematic review. *Educational Technology & Society,
10*(3), 247-256.


York: Scribner.


APPENDIX A: ESSENTIAL INFORMATION FOR SAMPLING DISTRIBUTIONS

- A sampling distribution of sample means (based on quantitative data) is a distribution of all possible sample means (statistics) for a given sample size randomly sampled from a population with mean $\mu$ and standard deviation $\sigma$. It is a probability distribution for the sample mean.
- The sampling distribution for means has the same mean as the population.
- As the sample size (n) gets larger, the variability of the sample means gets smaller (a statement, a visual recognition, and predicting what will happen or how the next picture will differ).
- Standard error of the mean is a measure of variability of sample statistic values.
- The building block of a sampling distribution is a sample statistic.
- Some values of statistics are more or less likely than others to be drawn from a particular population.
- The normal approximation applies in some situations but not others.
- If the normal approximation applies, then the empirical rule can be applied to make statements about how often the sample statistic will fall within, say, 2 standard deviations of the mean.
- Different sample sizes lead to different probabilities for the same statistic value (know how sample size affects the probability of different outcomes for a statistic).
- Sampling distributions tend to have the shape of a normal distribution rather than the shape of the population distribution, even for small samples.
- As sample sizes get very large, all sampling distributions for means look alike (i.e., have the same shape) regardless of the population from which they are drawn.
- Averages are more normal and less variable than individual observations.
- Be able to distinguish between a distribution of observations in one sample and a distribution of $x$ statistics (sample means) from many samples (sample size n greater than 1) that have been randomly selected.
APPENDIX B: LEARNING CONTENT

Ken: Hello. My name is Ken and I have a new job working on a shrimp farm. My boss is currently trying a food supplement to accelerate the growth of the shrimp. In order to see how this supplement works, she wants me to monitor the growth of the shrimp. In particular, she wants me to measure the average weight of shrimp in the pond weekly.

Q1. I know that it is impossible to measure every shrimp in the pond since there are more than 100,000 shrimp. So, I randomly draw samples of 30 shrimp each and compute the mean for each sample. The sample means are slightly different from each other. What should I do now?

Ask the expert

1. What is mean?

So we need to find out what a mean is. A mean is an arithmetic average of a defined set of numbers. What this means is that it is a value that is computed from all of the numbers that we have that defines pretty much where the middle of that group of numbers is. If we look at this equation, this means that we are going to add up all of the numbers; the summation of x. Then divide it by the total number of numbers that we have to give us a mean.

2. Why is each sample mean slightly different from each other?

Ok, so why are sample means slightly different from each other? To understand why this is the case, we need to understand what the process of sampling is. When we have a large population of data, say with 100,000 numbers or 100,000 data points in it, and we draw random samples. That means that any value in that population has an equal chance of being pulled out to be included in the sample. So if we draw a sample of 30 numbers, it can be any randomly chosen 30 values from that population, so if we draw a sample of 30 values and calculate a mean, we have a mean of this sample. If we then put all those numbers back and draw and new sample of 30 more values, it can be any other 30 randomly chosen values, this means that the 2 means calculated from these 2 samples might be slightly different.
3. How could we integrate the data we have into something meaningful?

Now let’s look at how to integrate this information that we have learned. Because there are far too many values in a population to get a direct measure of something of interest like a mean, we have to estimate this by taking a variety of samples and then generating an estimate of what we want to know but can’t directly measure. To understand how this works we need to talk about something called the central limit theorem. This is an idea that says if we draw a sample from a population, calculate a mean, and save that mean, then repeat this process many times, we can create what is called a sampling distribution of the means of the samples that we took. If we then take the mean of this sampling distribution of means, we will have a very close estimate of the true population mean.

*Run the simulation*

1. Show me how to do it.
   a. This simulation is to show you how a mean of distribution of sample means will be approximately equal to the mean of the population. First, we set the sample size to 30 and calculate one sample mean. Here you can see 30 cubes are randomly drawn from the population. They are put back to the population after calculating the mean. Click the save button to store this information.
   b. To speed up the process, let’s calculate 5 sample means and save the information.
   c. Repeat this process 3 times. There you can see a distribution gradually forming.
   d. Calculate 100 sample means. You can see the distribution of sample means become a normal distribution. That is, the skewness and kurtosis are close to 0. Click save button.
   e. Calculate 100 sample means again and save information.
   f. Calculate 1000 sample means and save information.
   g. Click show button to see the saved information. As you can see, the values in the mean column get closer to the population mean, 60.25, when the numbers of samples are larger.
h. Click Show Mean Graphic button and a diagram pops out. The X axis is the number of sample means; the Y axis is the sample means. The red bar is the mean of population, which is 60.25. As you can see, when the number of sample means increases, the mean of distribution of sample means gets closer to the red bar.

2. Let me try it.

Advice the client

1. Get rid of those which are too extreme and compute a mean of those which are closer.

2. Calculate the median of these sample means.

3. Calculate the mean of all sample means drawn (correct).

Q2. Suppose I choose to draw only one single sample and compute the mean. What could I do to make this single sample mean a more accurate estimate of the actual population mean?

Ask the expert

What is the effect of sample size?

Ok, let’s talk about the effect of sample size on the precision of our estimate of something from the population. To do this, I’ll illustrate based on what we can consider a population distribution of what we’re interested in. In Ken’s case, it’s a population distribution of the weight of the shrimp in the pond. If we take a sample of a very small number of values from this population, or in our case a very small number of shrimp from this population, let’s say three of them, we would take a sample of three. We could possibly get three values from the very high end of the distribution, or three values from the very low end of the distribution. The means of these two samples are going to be very high or very low, but not very representative of the majority of the actual values in this distribution. However, if we take a sample of, let’s say, 100 values from this distribution, it’s much more likely that those 100 values would span different parts of the distribution. This will reduce the likelihood that our sample mean is
either very low or very high. Therefore, increasing sample size tends to produce better estimates of the population value that we are interested in.

**Run the simulation**

1. Show me how to do it.
   
a. This simulation is going to show you how different sample sizes influence the estimate of the population mean. First, type sample size 2 and press “Enter” key.

b. Type sample size 100 and press “Enter” key.

c. Click x 1 button and you can see a sample of 2 scores is randomly drawn from the population in distribution 2. A sample of 100 scores is randomly drawn from the population in distribution 2. You can see that the mean of a sample with 100 scores is closer to the mean of the population, compared with the mean of the sample with 2 scores. But it is still possible that these two means are close to the mean of the population by chance. Let’s compute 100 sample means.

d. Click x 100 button to compute 100 sample means and observe the spreads of the sample means in distribution 1 and distribution 2. You can see that the sample means in distribution 2 is less spread out, compared with those in distribution 1. Less spread out means that the sample means are more precise to predict the mean of the population.

We can conclude that as the sample size increases, precision in estimate the population also increases.

2. Let me try it.

**Advice the client**

1. Decrease the sample size.

2. **Increase the sample size** (correct).
Q3. I decide to continue to take multiple samples from the pond. The means of the samples will not be identical and I would like some understanding of this variation. In particular, what will influence the variability of my sample means?

**Ask the expert**

1. What is variance? What is standard deviation?

   Well, remember we already talked about one descriptive characteristic of a group of data and that was the mean for the arithmetic center of that data. However, there’s another important characteristic of a distribution of data that we should mention; and that’s dispersion or spread of the data. Variance and Standard Deviations are measures of this and essentially what they speak to is the average distance of each value from the mean. In a distribution that has very high variance, it’s going to look very flat and spread out. That means on average each value is much further from the mean than in a distribution with low variance where it looks more tall and squished together. In these distributions, the average distance of each value from the mean is very small.

2. Why bother knowing the variation of data?

   What is the importance of knowing the variation of a distribution of data? Well the importance is knowing essentially what the data looks like. Are there a lot of extremely high and extremely low values that make the distribution very spread out? Or are most of the values very closely clustered around the mean that make the distribution look tall and squished together? Knowing the variation can help us understand something about the characteristics of that distribution of data.

3. What will influence the variability of my sample means?

   For example, let’s use sample size to illustrate this point. For one distribution we’ll use a very small sample size, let’s say two values. For another distribution, let’s take a larger sample size like 100 values. If we then calculate 100 sample means, we can see that there is a very big difference in the spread of the distribution for the small sample size and the large sample size. Specifically, the spread of distribution 2 with the large sample size is much less than the spread of distribution 1. Therefore we can conclude that as sample size increases, the
variability or spread of a distribution becomes smaller. There is a relationship between sample size and the spread or variability of a distribution.

**Run the simulation**

1. Show me how to do it.
   a. This simulation is going to show you how different sample sizes influence the variability of sample means. First, type 2 and press “Enter” key to set sample size as 2.
   b. Type 100 and press “Enter” key.
   c. Compute 5 sample means by clicking x 5 button. Observe the Standard Deviation (SD) in the population, distribution 1, and distribution 2. Repeat this process for several times. You will find that the sample means in distribution 1 are more dispersed than those in distribution 2.

2. Let me try it.

**Advice the client**

1. **The size of each sample determines the variability of the sampling distribution (correct).**

2. The size of each sample doesn’t influence the variability of the sampling distribution.

3. Small samples tend to underestimate the variability of the sampling distribution.

**Q4. I can measure variation with either the variance or standard deviation. If I know the variance (or standard deviation) of the weight of a shrimp for the entire population, can I predict the variance (or standard deviation) of the sampling distribution of shrimp weight?**

**Ask the expert**

1. What is variance? What is standard deviation?

   Well remember we already talked about one descriptive characteristic of a group of data and that was the mean for the arithmetic center of that data. However, there’s another important characteristic of a distribution of data that we should mention; and that’s dispersion or spread of the data. Variance and Standard Deviations are measures of this and essentially what they speak to is the
average distance of each value from the mean. In a distribution that has very high variance, it’s going to look very flat and spread out. That means on average each value is much further from the mean than in a distribution with low variance where it looks more tall and squished together. In these distributions, the average distance of each value from the mean is very small.

2. What is the relationship between population parameters and sample statistics?

Let’s talk about the relationship between populations and samples. A population is a group of data that you want to know something about. This could be the population of people in the United States, or in our case, the population of shrimp in the pond.

Typically, we cannot study populations directly because they are too large and it costs too much to make that type of study practical. However, we can draw samples and measure what we’re interested in the sample to infer something about the population.

When we talk about the characteristics of a population, they are always referred to as population parameters. These can be things like the mean, represented by the Greek symbol Mu, or the variance, represented by the symbol Sigma squared. Samples always have statistics. A sample mean is represented by the Roman character X with a bar over the top and a sample variance is represented by the Roman character S squared.

3. How are measures of variability related between samples and population?

In our simulation, let’s use an example of setting the sample size to 25 and then taking 100 sample means and plotting them in a distribution. Ok, we can see this distribution here. To see it in another form let’s save it and click show to look at a tabled form of this data. Ok, so we have our first distribution of 100 sample means. Let’s do this multiple times and look what happens to the Variance.

In the table, we can see that the Variance of these samples is very small compared to the actual Variance in the population distribution. However, the Variance of the samples seems to be related to each other. They are very small but they’re all around a similar value. As we said before, the Variance estimate is a
biased statistic. But if we divide by the sample size, we can very easily show that the Variance would then be a good estimation of the population Variance.

*Run the simulation*

1. Show me how to do it.
   a. This simulation is going to show you how to predict the standard deviation of population by using sampling distribution. First, click 30 button to set sample size as 30.
   b. Click x 5 button to compute 5 sample means and click save button to store information.
   c. Repeat this process for another 3 times. Click x 100 button and click save button.
   d. Repeat this process for another 3 times. Click Show button to see the stored button. As you can see, when the number of samples increases, the standard deviation of each sample gradually decreases and fluctuates around 0.91.
   e. Click Show SD Graphic button. You can see the white dots (the standard deviation of each sample) are closer to the red bar (the standard deviation of the population) when the number of the sample increases.

Why are these standard deviations of distribution of sample means close to 0.91? Based on Central Limit Theorem, the standard deviation of distribution of sample means is the standard deviation of the population divided by square root of the sample size. That is, 0.91 comes from the equation that 5 is divided by square root of 30.

2. Let me try it.

*Advice the client*

1. Yes. The variance of distribution of sample means is the variance of the population divided by the sample size (correct).
2. Yes. The variance of distribution of sample means is the variance of the population.
3. No. It is impossible to predict the variance of distribution of sample means from the variance of the weight of a shrimp for the entire population.
Q5. What will happen to the variation if I use samples of 5 shrimp instead of 25?

*Ask the expert*

How are measures of variability related between samples and population? In our simulation, let’s use an example of setting the sample size to 25 and then taking 100 sample means and plotting them in a distribution. Ok, we can see this distribution here. To see it in another form let’s save it and click show to look at a tabled form of this data. Ok, so we have our first distribution of 100 sample means. Let’s do this multiple times and look what happens to the Variance.

In the table, we can see that the Variance of these samples is very small compared to the actual Variance in the population distribution. However, the Variance of the samples seems to be related to each other. They are very small but they’re all around a similar value. As we said before, the Variance estimate is a biased statistic. But if we divide by the sample size, we can very easily show that the Variance would then be a good estimation of the population Variance.

*Run the simulation*

1. Show me how to do it.
   a. This simulation is going to show you how different sample sizes may influence the standard deviation of the distribution of sample means. First, type 5 and press “Enter” key to set sample size as 5.
   b. Type 25 and press “Enter” key.
   c. Compute 100 sample means by clicking x 100 button. Observe the Standard Deviation (SD) in distribution 1 and distribution 2. Repeat this process for several times. You will find that the sample means in distribution 1 are more spread-out than those in distribution 2.

2. Let me try it.

*Advice the client*

1. The variance of the distribution of sample means will decrease.

2. **The variance of the distribution of sample means will increase (correct).**

3. The variance of the distribution of sample means will not be influenced.
Q6. I am wondering what would happen if the shape of the distribution of shrimp weights were changed or even quite unusual. That is, will the shape of the distribution of shrimp weight in the population determine the shape of the distribution of sample means?

Ask the expert

1. How to describe distributions of data?

An instructor conducted a survey to investigate the students’ anxiety in an introductory statistics class during the first week of the class. One question asked was, “How stressed have you been in the last 1 week, on a scale of 0 to 10, with 0 being not at all stressed and 10 being as stressed as possible?” The 80 students’ responses were as follows:

0, 5, 8, 6, 7, 5, 5, 1, 7, 2, 4, 7, 9, 6, 3, 5, 1, 2, 8, 3, 6, 9, 6, 6, 3, 4, 6, 5, 6, 10, 7, 7, 2, 4, 4, 2, 10, 8, 7, 7, 3, 8, 7, 0, 4, 7, 3, 6, 0, 5, 5, 6, 8, 3, 3, 1, 5, 6, 9, 5, 1, 9, 6, 8, 4, 4, 4, 2, 8, 5, 0, 2, 1, 7, 4, 3, 9, 4

In its raw form, this data is very difficult to interpret. One solution is to make a table showing how many students used each of the 11 values. A frequency table makes the pattern of numbers easy to see. This displays the values (0 to 10), how many students responded with each value, and the percentage of all of the students who responded with each value. You can see most of the students rated their stress around 4 to 7.

<table>
<thead>
<tr>
<th>Frequency Table</th>
</tr>
</thead>
<tbody>
<tr>
<td>Valid</td>
</tr>
<tr>
<td>-------</td>
</tr>
<tr>
<td>0</td>
</tr>
<tr>
<td>1</td>
</tr>
<tr>
<td>2</td>
</tr>
<tr>
<td>3</td>
</tr>
<tr>
<td>4</td>
</tr>
<tr>
<td>5</td>
</tr>
<tr>
<td>6</td>
</tr>
<tr>
<td>7</td>
</tr>
<tr>
<td>8</td>
</tr>
<tr>
<td>9</td>
</tr>
<tr>
<td>10</td>
</tr>
<tr>
<td>Total</td>
</tr>
</tbody>
</table>
We can be sure that we have represented all of the data by looking at the very bottom. 80 responses are in the table that account for 100% of the data. Let’s try another good way of looking at this data but with a graph instead of a table. A graph can make a large group of scores easy to understand just by looking at it. This type of graph is called a histogram and it may look familiar, like a bar chart. Histograms represent the data values along the horizontal, or x-axis, and the number of times that value occurred along the vertical, or y-axis. This is the same data that we had in the frequency table, just shown with a picture. Notice again that most of the data is between the values of 4 and 7 represented by the tallest bars.

A frequency table or histogram tell us that how the data are spread out or distributed. Describing the shape of the distribution is very important in descriptive and inferential statistics. One critical aspect of a distribution’s shape is whether it has only one main high point. In our stress example, this highest point is 6. This is called a unimodal distribution. If a distribution has two fairly equal high points, it is called a bimodal distribution.
If all the values have the same frequency, it is called a uniform distribution.

A distribution that clearly is not symmetrical is called a skewed distribution. A skewed distribution has one side that is long and spread out, like a tail. The tail indicates the direction of the skew. If the tail is where higher values are, we call it a positively skewed distribution. If the tail is where lower values are, we call it a negatively skewed distribution. This distribution is positively skewed because the tail is where the higher values are.
Skewness can actually be assessed as a value. A distribution with skewness of 0 is a relatively normal distribution. Positive skewness values indicate a positive skew, and negative skewness values indicate a negative skew.

2. What is a normal distribution?

A normal distribution is a unimodal, symmetrical, and bell-shaped distribution. It has two important properties. First, the mean, the sum of all values divided by the number of the values, determines the center of the distribution.

The problem is that symmetry is not the only one condition because a symmetric distribution can be too flat or steep.
Kurtosis is a measure of whether the data are peaked or flat relative to a normal distribution. A value of 0 means the distribution is normal. So, we need the second property, which is that the standard deviation determines the shape of the distribution. A nice property of normal distributions is that we can get an idea of the responses using the mean, standard deviation, and the 68-95-99.7 rule. That is, 68% of the observations fall within one standard deviation of the mean. 95% of the observations fall within two standard deviations of the mean. 99.7 of the observations fall within three standard deviations of the mean.

3. Will the shape of the population influence the shape of distribution of sample means?

If we begin with a normally distributed population and set a sample size very low, such as 2, and compute 1000 sample means we can plot this new distribution of sample means. The shape of the population distribution is relatively normal and we can also see that the shape of the sampling distribution is relatively normal.

Let’s take a look at a Skewed Distribution. In a skewed distribution, the shape of the population is not normal and the shape of the distribution of sample means is also not normal. Remember, we are doing this with very small sample sizes. For a uniform distribution, the shape of the population distribution is not normal and the shape of the distribution of sample means is also not normal. For a Bimodal Distribution, the shape of the population distribution is not normal and the shape of the sample distribution is also not normal. Again, in a U-shaped
Distribution, the population distribution is not normal and the shape of the sampling distribution is also not normal. So to summarize, with a very small sample size, such as a sample size of 2 and many sample means, the shape of the population seems to determine the shape of the sampling distribution. With a normal population distribution, we get a normal sampling distribution. With abnormal population distributions, we tend to get abnormal sampling distributions.

However, let’s increase the sample size to 30. So we are going to make the sample size 30 and take many samples; 1000 samples. We can see here that the population distribution is normal and the distribution of the sample means is also normal.

Here is an abnormal population distribution, a skewed distribution. But what do we see? The sampling distribution is normal. Here is the uniform population distribution, an abnormal distribution, but the sampling distribution is normal. Here is the Bimodal distribution, which is not normal in the population, but then again, the sampling distribution is normal. The U-shaped distribution, not normal in the population, but the sampling distribution is normal. So if we come to a conclusion from this data; the shape of the population determines the shape of the sampling distribution only when the sampling size is less than 30. When the sampling size becomes greater than 30, we tend to see that the sampling distribution is normal regardless of the shape of the population.

Run the simulation

1. Show me how to do it.
   a. This simulation is going to show you the relationship between the shape of the population and the distribution of sample means. First, click 2 button to set sample size as 2.
   b. Click x 1000 button to compute 1000 sample means. Look at the shapes of both distributions. The population is normal; the shape of distribution of sample means is approximately normal.
   c. Let’s see different population. Click reset button and then click change distribution button to switch to skewed distribution. Set sample size as 2.
by clicking 2 button. Click x 1000 button to compute 1000 sample means. Look at the shapes of both distributions. The population is not normal, so does the shape of distribution of sample means.

d. How about uniform distribution? Let’s click reset button and then click change distribution button 2 times to switch to uniform distribution. Set sample size as 2 by clicking 2 button. Click x 1000 button to compute 1000 sample means. Look at the shapes of both distributions. The population is not normal, so does the shape of distribution of sample means.

e. How about bimodal distribution? Let’s click reset button and then click change distribution button 3 times to switch to bimodal distribution. Set sample size as 2 by clicking 2 button. Click x 1000 button to compute 1000 sample means. The population is not normal, so does the shape of distribution of sample means.

f. How about U-shape distribution? Let’s click reset button and then click change distribution button 4 times to switch to U-shape distribution. Set sample size as 2 by clicking 2 button. Click x 1000 button to compute 1000 sample means. The population is not normal, so does the shape of distribution of sample means.

g. Let’s summarize a little bit. With sample size of 2 and computing 1000 sample means from randomly drawing, the shape of population seems to determine the shape of distribution of sample means. That is, when the shape of population is normal, so does the shape of distribution of sample means. When the shape of population is not normal, the shape of distribution of sample means is not normal as well. Let’s try the sample size of 30.

h. Set sample size as 30 by clicking 30 button. Click x 1000 button to compute 1000 sample means. The population is normally distributed, so does the shape of distribution of sample means.

i. Let’s see different population. Click reset button and then click change distribution button to switch to skewed distribution. Set sample size as 30
by clicking 30 button. Click x 1000 button to compute 1000 sample means. The population is not normally distributed, but the shape of distribution of sample means is approximately normal.

j. How about uniform distribution? Let’s click reset button and then click change distribution button 2 times to switch to uniform distribution. Set sample size as 30 by clicking 30 button. Click x 1000 button to compute 1000 sample means. The population is not normal, but the shape of distribution of sample means is approximately normal.

k. How about bimodal distribution? Let’s click reset button and then click change distribution button 3 times to switch to bimodal distribution. Set sample size as 30 by clicking 30 button. Click x 1000 button to compute 1000 sample means. The population is not normal, but the shape of distribution of sample means is approximately normal.

l. How about U-shape distribution? Let’s click reset button and then click change distribution button 4 times to switch to U-shape distribution. Set sample size as 30 by clicking 30 button. Click x 1000 button to compute 1000 sample means. The population is not normal, but the shape of distribution of sample means is approximately normal.

m. Let’s come to a conclusion. The shape of the population determines the shape of the distribution of sample means (if the shape of the population is normally distributed, so does the shape of distribution of sample means), but the effect gets smaller as the sample size increases and is relatively unimportant when sample size is larger than 30.

2. Let me try it.

*Advice the client*

1. No, the shape of the distribution of sample means is not determined by the shape of the distribution of shrimp weight in the population.

2. Yes, the shape of the distribution of sample means is determined by the shape of the distribution of shrimp weight in the population and not by the sample size.

3. **Yes, but the effect gets smaller as the sample size increases and is relatively unimportant for N>30 (correct).**
APPENDIX C: FINDINGS FROM THE SECOND PILOT STUDY

1. I was unclear as to whether each of the red buttons had to be completed. For example, some of the simulations were the same for multiple questions. Was I supposed to re-watch the simulations? I wasn't sure.

2. More amicable/younger tutors. I couldn’t really relate to the actors in a way that I became interested in the material wholly.

3. I only got to Question 6 in the tutorial so maybe you should get more time. I understood the tutorial but I got confused on the questions in the post question assignment.

4. I would have liked a little bit more time to fully utilize the "Ask the expert" option. Also, labels on the graphs would be helpful in understanding the simulations.

5. An illustrated user guide will helpful.

6. Increase time and provide post-test type questions in the tutorial

7. More explanation of what is required and what is not required before moving on to "Advise Ken".
APPENDIX D: REPLY MESSAGE FROM DR. SCHANK

You have the basic idea but you need to start with a really simply issues and let students practice that and so on; nothing can be taught in 30 minutes.

Roger Schank
http://www.rogerschank.com/

On Feb 24, 2009, at 5:18 PM, Jack wrote:
Hey Dr. Schank,
Here is a link to the screenshots of my tutorial. Thank you very much.
http://oak.cats.ohiou.edu/~ch243905/Intro/Intro.html
Best regards,
Jack

Roger Schank wrote:
Send me a one pager
roger schank
http://www.rogerschank.com/

On Feb 23, 2009, at 5:33 PM, Jack wrote:
Hey Dr. Schank,
I am a doctoral student of Instructional Technology at Ohio University. I developed a computer simulation based on Goal-Based Scenarios to teach statistical concepts, sampling distributions. I was wondering you can take a minute of your time to look at my tutorial or have one of your colleagues or students evaluate it. If you or anyone else is interested, I will send my tutorial. Thank you very much. I greatly appreciate your time.
Sincerely,
Jack
APPENDIX E: PRETEST ITEMS

1. The central limit theorem states that the distribution of sample means is normal only when the population is normal.
   A. True -----------------------B. False

   How confident are you that you chose the correct answer? (Circle one of the values below.)
   20% 30% 40% 50% 60% 70% 80% 90% 100%

2. When the sample size is more than 30, the effect that the shape of the population has on the shape of the distribution of sample means is relatively unimportant.
   A. True -----------------------B. False

   How confident are you that you chose the correct answer? (Circle one of the values below.)
   20% 30% 40% 50% 60% 70% 80% 90% 100%

3. The distribution of sample means becomes approximately normal only when the standard deviation of the population is large.
   A. True -----------------------B. False

   How confident are you that you chose the correct answer? (Circle one of the values below.)
   20% 30% 40% 50% 60% 70% 80% 90% 100%
4. The histogram below is a distribution for a population with mean 60 and standard deviation 11. If one randomly draws 500 samples of size 30 from this population, computes the mean of each sample and plots the sample means, which histogram is the distribution of sample means?

A  

B  

C  

D  

How confident are you that you chose the correct answer? (Circle one of the values below.)

20% 30% 40% 50% 60% 70% 80% 90% 100%
5. The histogram below is a distribution for a population with **mean 55** and **standard deviation 6.27**. If one randomly draws 500 samples of **size 35** from this population, computes the mean of each sample and plots the sample means, which histogram is the distribution of sample means?

**How confident are you that you chose the correct answer? (Circle one of the values below.)**

20% 30% 40% 50% 60% 70% 80% 90% 100%
6. The histogram below is a distribution for a population with **mean 55** and **standard deviation 6.27**. If one randomly draws 500 samples of **size 2** from this population, computes the mean of each sample and plots the sample means, which histogram is the distribution of sample means?

![Population Distribution](image)

**A.** Distribution of sample

![Distribution of sample](image)

**B.** Distribution of sample

![Distribution of sample](image)

**C.** Distribution of sample

![Distribution of sample](image)

**D.** Distribution of sample

*How confident are you that you chose the correct answer? (Circle one of the values below.)*

20% 30% 40% 50% 60% 70% 80% 90% 100%
7. One randomly draws many samples of size 50 from a population. The histogram below is a distribution of sample means with mean 60. If all the four populations below have the same mean of 60, which histogram is the population that the distribution of sample means has been drawn from?

D. All these four populations are possible.

How confident are you that you chose the correct answer? (Circle one of the values below.)

20%  30%  40%  50%  60%  70%  80%  90%  100%
8. One randomly draws 500 samples with sample size of 5 and 500 samples with sample size of 35 from the same population, computes the mean of each sample and plots the sample means. The figures below are the distributions of the sample means for each sample size. Which histogram is the one with the sample size of 35?

A. Distribution of sample

B. Distribution of sample

How confident are you that you chose the correct answer? (Circle one of the values below.)

20%   30%   40%   50%   60%   70%   80%   90%   100%
APPENDIX F: POSTTEST ITEMS

1. One randomly draws 500 samples with **sample size of 5** and 500 samples with **sample size of 35** from the same population, computes the mean of each sample and plots the sample means. The figures below are the distributions of the sample means for each sample size. Which histogram is the one with **the sample size of 5**?

   A. Distribution of sample

   B. Distribution of sample

   *How confident are you that you chose the correct answer? (Circle one of the values below.)*

   20% 30% 40% 50% 60% 70% 80% 90% 100%

2. The central limit theorem states that the distribution of sample means is normal only when the population is normal.

   A. True -----------------------B. False

   *How confident are you that you chose the correct answer? (Circle one of the values below.)*

   20% 30% 40% 50% 60% 70% 80% 90% 100%

3. The distribution of sample means becomes approximately normal only when the standard deviation of the population is large.

   A. True -----------------------B. False

   *How confident are you that you chose the correct answer? (Circle one of the values below.)*

   20% 30% 40% 50% 60% 70% 80% 90% 100%
4. When the sample size is more than 30, the effect that the shape of the population has on the shape of the distribution of sample means is relatively unimportant.
   
   A. True  -----------------------B. False

   How confident are you that you chose the correct answer? (Circle one of the values below.)
   
   20%   30%   40%   50%   60%   70%   80%   90%   100%
5. The histogram below is a distribution for a population with \textbf{mean 55} and \textbf{standard deviation 6.27}. If one randomly draws 500 samples of \textbf{size 35} from this population, computes the mean of each sample and plots the sample means, which histogram is the distribution of sample means?

\begin{itemize}
\item[A.] Population
\item[B.] Distribution of sample
\item[C.] Distribution of sample
\item[D.] Distribution of sample
\end{itemize}

\emph{How confident are you that you chose the correct answer? (Circle one of the values below.)}

20% 30% 40% 50% 60% 70% 80% 90% 100%
6. The histogram below is a distribution for a population with **mean 55** and **standard deviation 6.27**. If one randomly draws 500 samples of **size 2** from this population, computes the mean of each sample and plots the sample means, which histogram is the distribution of sample means?

![Population Distribution](image)

![Distribution of sample A](image)

![Distribution of sample B](image)

![Distribution of sample C](image)

![Distribution of sample D](image)

*How confident are you that you chose the correct answer? (Circle one of the values below.)*

- 20%
- 30%
- 40%
- 50%
- 60%
- 70%
- 80%
- 90%
- 100%
7. One randomly draws many samples of size 50 from a population. The histogram below is a distribution of sample means with mean 60. If all the four populations below have the same mean of 60, which histogram is the population that the distribution of sample means has been drawn from?

A. 

B. 

C. 

D. All these four populations are possible.

How confident are you that you chose the correct answer? (Circle one of the values below.)

20% 30% 40% 50% 60% 70% 80% 90% 100%
8. The histogram below is a distribution for a population with **mean 60** and **standard deviation 11**. If one randomly draws 500 samples of **size 30** from this population, computes the mean of each sample and plots the sample means, which histogram is the distribution of sample means?

How confident are you that you chose the correct answer? (Circle one of the values below.)

20% 30% 40% 50% 60% 70% 80% 90% 100%
APPENDIX G: INTERVIEW QUESTION

- Could you tell me the characteristics of sampling distribution?
- Could you tell me about your experience working with this tutorial? (elaborate)
- Does this tutorial help you learn statistics concepts? Could you give me an example?

**Goal**
- You know the goal of this tutorial is to help students learn the statistical concept.
- How do you feel about learning statistics? Is it interesting to you, or not?
- When Ken asked you questions, did you generate inquiries in your mind?
- What do you think about these inquiries?

**Mission**
- You are playing as a statistics specialist, right? So, do you know your mission in this tutorial?
- What do you think about the mission?
- Is the mission motivational for you to pursue?

**Cover story**
- How do you feel the cover story?

**Role**
- What do you think about the role you played in the tutorial?
- How does this role influence you?
- Does playing as a statistics specialist motivate you?

**Scenario operations**
- What do you think about the scenario operations? Here the scenario operations mean that Ken asks you question and you can choose to ask the experts and run the simulation.
- What did you think about having control over the resources?

**Recourses**
- How would you feel the recourses provided in the tutorial?
- Which one do you prefer? Why?
- Is it sufficient for you to advise Ken?
• Feedback
  o What do you think about the feedback?
  o How did you feel when you fail to give Ken correct advice? Are you frustrated or motivated?
• What components help you best? Why?
• What components help you least? Why?
• What are the strengths?
• What are the weaknesses?
• Are you comfortable with this tutorial?
• Is the tutorial (instruction) distracted?
• Is the instruction accomplishable with reasonable effort?
• Is the instruction time-consuming?
• Is the information or instruction clear?
• How do you feel the sequence, the pace, and the transition in the tutorial?
• How do you feel the size of the content in each question?
• How would you redesign and improve the instruction if you have a chance to design it?
• Are there any suggestions or comments that you would like to make?

What do you think about your learning style? Are you more independent or more dependent?
APPENDIX H: RETENTION TEST ITEMS

1. What is the distribution of sampling means?

2. If one draws a sample of size 10 from the population below, calculates a mean, puts
the sample back to the population, and repeats these three steps for 1000 times, what
is the mean of these 1000 sample means?

Please write down your answer and explanation:
3. One randomly draws 500 samples with **sample size of 50** and 500 samples with **sample size of 2** from the same population, computes the mean of each sample and plots the sample means. The figures below are the distributions of the sample means for each sample size. Which histogram is the one with **the sample size of 50**?

Please write down your explanation:
4. One randomly draws many samples of **size 50** from a population. The histogram below is a distribution of sample means with **mean 60**. If all the three populations below have the same mean of 60, which histogram is the population that the distribution of sample means has been drawn from?

![Distribution of sample means](image)

- **A**
- **B**
- **C**
- **D** All these **populations** are possible.

**Please write down your explanation:**
APPENDIX I: IRB

The amendment, detailed below, and submitted for the following research study has been approved by the Institutional Review Board at Ohio University.

**Project:** Formative Research on a Computer Simulation for Teaching Statistics Concepts

**Amendment:** Utilize Qualitative Methodology; Revised Consent Form, Revised # of Participants, etc.

**Primary Investigator:** Chung-Yuan Hsu

**Co-Investigator(s):**

**Advisor:** David Moore

**Department:** Educational Studies

Rebecca G. Cale, AAB, CIP
Office of Research Compliance

Date: 3/05/09
APPENDIX J: STORYBOARD

Screen name: Statistics specialist
Screen No.: 1
Serial No.: 1

Screen description
1. A is a welcome page displaying graphics.
2. B displays the title of the tutorial.
3. C is a button taking to Screen No. 2 when clicked.

Screen layout

---

Screen name: Statement of learning goal
Screen No.: 2
Serial No.: 2

Screen description
1. A is the statement of learning goal.
2. B is a button taking to Screen No. 3 when clicked.

Screen layout

This tutorial will help you learn the features of the sampling distribution of the sample mean, a theorem related to the Central Limit Theorem (CLT).

Next
Screen name: Statement of mission

You are going to play a role as a statistics specialist.
Your first client is....

Enter

Screen layout

Screen name: Statement of cover story

Screen description
1. A provides scenario operations including running simulation, asking the expert, and advising Ken.
2. B displays six questions asked by Ken. A user can only review previous questions.
3. C is a video clip in which Ken prompts the question. On entering the screen, Ken begins introducing himself and describing his problems.
4. D is a control panel allowing a user to play, pause, and replay. The status bar below allows to selectively review certain sections of a video clip.

Screen layout

A

C

D

B
Screen name: User guide  

Screen description
1. A is a female guide, saying that “Hey, my name is Ashley. I’m here to help you with your activities as a statistics assistant. At this point, you can choose to run the simulation or talk to the expert first.” When the guide pops up, a black and translucent image covers the background except the left menu panel.
2. B includes two buttons. One will replay when clicked; the other will take to the screen 4 when selected.

Screen layout

Screen name: Run simulation  

Screen description
When entering this screen first time, a guide pops up, saying “Here you can choose to receive the instruction first or try the simulation first.”
1. A allows users to choose different activities.
2. B allows users to review previous questions.
3. C shows two options. “Show me how to do it” button links to a worked example (see Screen 4-2-1). “Let me try it” takes to the simulation (see Screen 4-2-2).
Screen name: Show me how to do it

Screen layout

A  C

B  D

Screen description
1. Allows users to choose different activities.
2. Allows users to review previous questions.
3. Is a worked example showing how to run the simulation.
4. Is a control panel allowing a user to play, pause, and replay. The status bar below allows to selectively review certain sections of a video clip.

Screen name: Let me try it

Screen layout

A  C  D

B  E

Screen description
1. Displays text instruction step by step showing how to run the simulation.
2. Goes back the menu (see Screen 4-2).
3. Consists of a button that changes different populations and information about the population and distributions of sampling means.
4. Shows graphics of population and distribution of sampling means.
5. Provides parameters for learners to manipulate distribution of sampling means.
Screen name: Ask the expert menu

Screen layout

A

C

B

Screen description
When entering this screen first time, a guide pops up, saying "Here you can select three questions to ask that expert".
1. A allows users to choose different activities.
2. B allows users to review previous questions.
3. C provides questions for users to ask the expert, including both fundamental and critical concepts.

Q1 Q2 Q3 Q4 Q5

Screen name: Ask the expert

Screen layout

A

C

B

D

E

Screen description
1. A allows users to choose different activities.
2. B allows users to review previous questions.
3. C displays a video clip in which an expert gives explanation for the question selected.
4. D links to the menu page (see Screen 4-3) when selected.
5. E is a control panel allowing a user to play, pause, and replay. The status bar below allows to selectively review certain sections of a video clip.
Screen name: Advise Ken

Screen description
When entering this screen first time, a guide pops up, saying “Good job! Look like you are ready to give Ken some advice. All you need to do is to pick the best answer for Ken. If you would like to view the question again, just click the play button”.
1. A allows users to choose different activities.
2. B allows users to review previous questions.
3. C gets rid of those which are too extreme and compute a mean of those which are closer.
4. D allows users to view the question when selected.
5. E calculates the median of these sample means.
6. F calculates the mean of all sample means drawn.

Screen layout
A  C
B  D

Screen name: Feedback

Screen description
1. A allows users to choose different activities.
2. B allows users to review previous questions.
3. C is a video clip in which the guide Ashley offers feedback. If an accurate advice is given, she says “Nicely done! That is a good explanation”. In contrast, she says “No, it is not right. You may need to go back to ask the expert or run the simulation again” if the user offers a wrong advice.
4. D is a button taking to Screen No. 2 when clicked.
Dear Specialist,
Thanks for your advice. Now I can monitor the growth of the shrimp population. With accurate measurements, we can figure out the best way to grow big shrimp. I appreciate your help.

Best regards,
Ken

Screen description
1. A is disabled.
2. B allows to review previous questions.
3. C displays a gratitude letter from Ken.
APPENDIX K: SCREENSHOTS OF STATISTICS SPECIALIST

Main page

This tutorial will help you learn the features of the sampling distribution of the sample mean, a theorem related to the Central Limit Theorem (CLT).

Statement of learning goal
Cover story

User guide
Worked example

This simulation is to show how a mean of distribution of sample means will be approximately equal to the population mean.

1) We set the sample size to 30 and calculate 1 sample mean.

Here you can see 30 cubes are randomly drawn from the population. They are put back to the population after calculating the mean.

2) Click the save button to store this information.

3) To speed up the process, lets calculate 5 sample means and save the information.

4) Repeat Step 3 for another 3 times.

There you can see a distribution gradually forming.

5) Calculate 100 sample means. Click save button.

You can see the distribution of sample means becomes a normal distribution. That is, the skewness and kurtosis are close to 0.

6) Calculate 500 sample means again and save information.

7) Calculate 1000 sample means and save information.

8) Click save button to see the saved information.

Running the simulation
Ask the expert

What is mean?

Why is each sample mean slightly different from each other?

How could we integrate the data we have into something meaningful?

Questions to ask the expert

The expert’s explanation of the concept
Advising the client

Dear Specialist,

Thanks for your advice. Now I can monitor the growth of the shrimp population. With accurate measurements, we can figure out the best way to grow big shrimp. I appreciate your help.

Best regards,

Ken

Feedback from the client