Copyright

by

Pei-Yu Wang

2010

The Dissertation Committee for Pei-Yu Wang Certifies that this is the approved version of the following dissertation:

The Impact of Animation Interactivity on Novices' Learning of Introductory Statistics

Committee:

Min Liu, Supervisor

Brandon Vaughn

Daniel Robinson

Helen Taylor Martin

Paul Resta

The Impact of Animation Interactivity on Novices' Learning of Introductory Statistics

by

Pei-Yu Wang, B.Ed.; B.B.A.; M.S.

Dissertation

Presented to the Faculty of the Graduate School of

The University of Texas at Austin

in Partial Fulfillment

of the Requirements

for the Degree of

Doctor of Philosophy

The University of Texas at Austin

May 2010

Dedication

For my parents,

Gui-Yan Cai and Zhi-Jie Wang,

and their never-ending supports.

Acknowledgements

Completing this dissertation would not have been possible without the help and support of numerous individuals. First and foremost, I give my heartfelt thanks to Dr. Min Liu, my supervising professor. Her insightful comments on the dissertation and uncompromising high standards in research make this dissertation a study of which I can be proud. I am extremely thankful to the members of my committee. A special word of gratitude goes to Dr. Brandon Vaughn whose full support and unfailing enthusiasm guided me throughout this process. I thank Dr. Paul Resta, whose constructive feedbacks encouraged me to face and overcome difficulties. I also thank both Dr. Daniel Robinson and Dr. Taylor Martin for their guidance in shaping the design of my study. Meanwhile, I would like to extend my gratitude to my master advisor in Taiwan, Dr. Janet Mei-chuen Lin, for constantly encouraging me to succeed in professional goals.

Finally, I would like to thank my parents, whose love and support were essential for the completion of this study. Without them, it would have been so much harder to finish the degree and accomplish this dissertation.

The Impact of Animation Interactivity on Novices' Learning of Introductory Statistics

Publication No._____

Pei-Yu Wang, Ph.D.

The University of Texas at Austin, 2010

Supervisor: Min Liu

This study examined the impact of animation interactivity on novices' learning of introductory statistics. The interactive animation program used in this study was created with Adobe Flash following Mayer's multimedia design principles as well as Kristof and Satran's interactivity theory. The research was guided by three main questions: 1) Is there any difference in achievement improvement among students who use different interactive levels of an animation program? 2) Is there any difference in confidence improvement among students who use different interactive levels of an animation program? 3) Is there any difference in program perception among students who use different interactive levels of an animation program?

This study was a one-way design where the independent variable was animation interactivity. In addition to a control group (Static Group) provided with only static materials, there were three groups with different levels of animation interactivity: 1) Animation with simple interactivity (Simple Animation Group), 2) animation with input

manipulation (Input Group), and 3) animation with practice and feedback (Practice Group). A sample of 123 college students participated in the study and was randomly assigned into groups. They gathered in the computer lab to work with the animation program and then took online surveys and tests for evaluation. Students were expected to learn *Principles of Hypothesis Testing* (concepts of type I error, type II error and *p*-value). The data collected in this study included 1) student learning attitudes, 2) achievement and confidence pre-test scores, 3) achievement and confidence post-test scores, and 4) program perception. Also, student manipulation of the animation program was recorded as Web log data. The data were analyzed by using multivariate analysis (MANOVA), univariate analysis (ANOVA), regression analysis, regression tree analysis and case analysis.

The findings were as follows: 1) Animation interactivity impacted students' improvement in understanding (p=.006) and lower-level applying (p=.042), 2) animation interactivity did not impact student confidence and program perception, 3) the regression analysis indicated that student prior knowledge and interest were the most important predictors on student achievement post-test scores instead of program manipulation, and 4) the regression tree showed that there were interactions among student interest, prior knowledge, and program manipulation on the achievement post-test scores. The case analysis showed that not all students manipulated the interactive animation program as expected due to a lack of motivation and cognitive skills, and this could decrease the effect of the interactive animation. This study hoped to broaden theories on interactive learning and serve as a reference for future statistics curriculum designers and textbook publishers.

Table of Contents

List of Tables xi

List of Figures xii

Chapter 1: Introduction 1

Learning statistics: A story of struggle 1

Challenges of integrating technology into statistics classroom 3

Statement of problem 8

Research questions 9

Significance of the study 9

Chapter 2: Literature Review 11

Theoretical foundations of multimedia learning 11

Information-processing theory 12

Dual-coding theory 13

Abstract knowledge acquisition 14

Feedback theory 16

Instructional use of interactive visuals 18

Computer-based visual aids in education 19

Interactivity and learner individual difference 23

Learning statistics with interactive technology 31

Concept development in statistics 31

The use of interactive technology in statistics education 34

Cognitive evaluation models 38

A general model: Bloom's revised cognitive taxonomy 38

A specific model in statistics learning: Garfield's reasoning model 40

Summary 42

Chapter 3: Methodology 44

Design and development of an interactive animation program 44

Learning objectives 44
Program design and development 46
Research design and procedures 54
Research questions and hypotheses 55
Pilot study results and implications 56
Subjects 59
Treatment 60
Experimental procedure 62
Instruments and data collection 63
Data analysis 68
Summary 73
Chapter 4 Results 74
Participant background 74
Results of data analysis 74
Reliability of instrument 74
Pre-test data analysis 75
Post-test data analysis 79
Web log data analysis 89
Summary 116
Chapter 5 Discussion and Conclusion 117
Summary of the study 117
Discussion of the results 118
Question1: Animation interactivity and achievement improvement .118
Question2: Animation interactivity and confidence improvement 122
Question3: Animation interactivity and program perception 125
Web log data analysis 126
Implications 128
Theoretical implications 128
Methodological implications 129
Pedagogical implications 130

Limitations 131
Conclusions 132
Appendix A – The pilot study results 134
Appendix B – Survey of attitudes toward statistics (SATS-36) 135
Appendix C – Pre-test questions for the pilot study 138
Appendix D – Post-test questions for the pilot study 145
Appendix E – Program perception survey for the pilot study 153
Appendix F – Pre-test questions for the main study 154
Appendix G– Post-test questions for the main study 163
Appendix H – Program perception survey for the main study 172
Appendix I – Informed consent form for participants 173
Appendix J – Bonferroni correction for post hoc comparisons 175
Appendix K – Table for the effect size 176
References 177
Vita - Pei-Yu Wang 184

List of Tables

Table 2.1 Continuum of interactivity (based on Kristof & Satran, 1995; Teo, Oh, Liu, & Wei, 2003) 28
Table 2.2 Students' misunderstanding in different statistics units 33
Table 2.3 Mapping Garfield's reasoning model to Bloom's revised cognitive taxonomy 41
Table 3.1 Reliability coefficients for pilot study instruments 57
Table 3.2 Treatment 60
Table 3.3 The experimental procedure 62
Table 3.4 SATS-36 scores 64
Table 3.5 The organization of achievement test questions 66
Table 3.6 Achievement and confidence test scores 67
Table 3.7 Program perception scores 67
Table 3.8 Web log data 68
Table 3.9 Summary of data analysis 71
Table 4.1 Distribution of subjects 74
Table 4.2 Reliability coefficients for main study instruments 75
Table 4.3 Descriptive statistics on SATS-36 scores for each group 76
Table 4.4 Descriptive statistics on achievement and confidence pre-test scores for each group 77
Table 4.5 Total seconds spent for each group 79
Table 4.6 Descriptive statistics on achievement and confidence post-test scores for each group 79
Table 4.7 Descriptive statistics on achievement and confidence improvement for each group 81
Table 4.8 MANOVA on learning improvement 83
Table 4.9 Following-up ANOVAs on learning improvement 84
Table 4.10 Tukey HSD post-hoc comparisons on learning improvement 86
Table 4.11 Descriptive statistics on program perception for each group 87
Table 4.12 ANOVA on program perception 87
Table 4.13 Summary of hypotheses testing results 88
Table 4.14 R square of the regression analysis 90
Table 4.15 Significance level of the regression analysis 90
Table 4.16 Regression coefficients 90
Table 4.17 Case selection 95

List of Figures

Figure 3.1 Sample page of a traditional textbook for statistical language of hypothesis testing 45
Figure 3.2 Program interface for the Question Statement Page 47
Figure 3.3 Program interface for the Static Instruction Mode 48
Figure 3.4 Program interface for the Simple Animation Mode 50
Figure 3.5 Program interface for the Change Inputs Mode 51
Figure 3.6 Program interface for the Practice Mode 53
Figure 3.7 The experimental procedure of the pilot study 57
Figure 4.1 Regression tree: Individual differences and animation manipulation 92
Figure 5.1 Animation interactivity and knowledge levels 119

Chapter 1: Introduction

LEARNING STATISTICS: A STORY OF STRUGGLE

I had my first statistics class in 1997 when I was a college freshman in Taiwan. It was a fundamental course targeted toward lower level students in the college of education. The course had simple materials, an organized schedule and an enthusiastic professor. However, after this one-semester statistics course, I was still unable to set the null hypothesis and the alternative hypothesis to solve an applied question. I found that I could not correctly interpret these abstract concepts. Like most college students in Taiwan, I was well trained in computational and applied mathematics from rigorous high-school courses. I was supposed to learn college statistics well, just like I had learned mathematics, but I felt frustrated and scared about college statistics. One year later, I enrolled in another introductory-level statistics course in the business school. I joined this course in order to gain more statistical skills. Though I tried hard to make every effort in this course, my statistical confidence and ability did not improve. Instead, everything became worse. After this course, I was able to memorize more equations and plug in numbers for computation, but I still had no idea how and why statistical procedures were utilized. These two courses almost destroyed my interest and patience for statistics, and I gave up on learning statistics for almost ten years.

I came back to statistics learning at UT in 2007. This was my third time taking an introductory-level statistics course. Due to my past horrible experiences in learning statistics, I had low expectations for the course. I viewed it as no more than a doctoral prerequisite. However, the instructor's teaching method evoked my enthusiasm and changed my thoughts about statistics. Most instructors started by teaching descriptive statistics, but he did not. Instead, he began with the unit *Principles of Hypothesis Testing*.

Included in this unit were the use of the null hypothesis and alternative hypothesis, the recognition of the direction of extreme, the computing of type I error, type II error and the *p*-value, and even the power of the test. These had been the most difficult and abstract concepts for me in previous fundamental statistics courses.

Are novice learners capable of grasping the most abstract concepts at the earliest stages of statistics learning? When my instructor was asked this question, his answer was very affirmative. He said that he wanted students to clarify the most difficult concepts in the most energetic stage, which is the beginning of the semester. Besides, he designed the material with concrete examples and graphics, which were easier for beginners in order to adapt such abstract statistics concepts for novice learners. He usually had colorful balls in bags with him in the class, and asked students to make statistical decisions based on the balls drawn. The following is a typical question he would ask:

I have two bags: Bag A and Bag B. Bag A has 10 blue balls and 5 red balls. Bag B has 1 blue ball and 8 red balls. A bag is picked at random and a ball is drawn at random from that bag and shown to you. Based on the color of the ball, you must decide if the ball was picked from Bag A or Bag B. Consider the following hypothesis:

The null hypothesis is: The ball was picked from Bag A

The alternate hypothesis is: The ball was picked from Bag B

Also consider the following decision rule: Reject H0 if the ball is red. Now, can you compute α*, the type I error?*

He asked students to first clarify the definition of the type I and type II error, and then look for clues in the question for computation. The type I and type II error problems are difficult for most novice learners. But with the concrete examples of "bags," "boxes," or "envelopes," statistics was no longer as difficult for me. This particular teaching method is valuable for use in all introductory-level classes.

I believe this struggle with learning statistics is not limited to myself, but is common to many other students. As a result, I created an online animated program based on my instructor's teaching method in order to help learners who also struggle with learning type I error, type II error and the *p*-value. This online animation was created to popularize this particular teaching method. To benefit learners with different backgrounds and preferences, this program was designed with different interactive levels for learners to play with. Also, with the use of the Internet, both on-campus and off-campus learners can take advantage of it.

After the online animation was created, some questions occurred to me. For example, does this animation work as well as physical teaching? How does this animation impact student learning? How do students interact with this animation? Does the animation meet the existing design principles? All of these questions motivated the initiation of this dissertation.

This study aimed to contribute to the improvement of statistics education. The main goals for this dissertation included designing and developing an online animated program based on this particular teaching method and understanding how students interact with this program to learn introductory statistics. Before delving into the design of the program and exploring the effects of using this program, it is necessary to discuss both the necessity and challenges of integrating technology into statistics learning.

CHALLENGES OF INTEGRATING TECHNOLOGY INTO STATISTICS CLASSROOM

The teaching and learning of statistics has pervaded all levels of education, including post-secondary, college-level and graduate-level curricula (Garfield & Ahlgren, 1988). In the United States, there is a growing trend to introduce concepts of statistics and probability into schools. Several important projects have been initiated, such as the Quantitative Literacy Project by the American Statistical Association (Landewehr, Swift,

& Watkins, 1987), Beyond Eighth Grade by the National Center for Research on Vocational Education and Campaign for Economic Literacy by the National Council on Economic Education (Steen, 1999). In addition to these research projects, the establishment of a national curriculum standard in statistics in 1989 also greatly popularized statistics education at the pre-college level (National council of teachers of Mathematics, 1989).

For college- and graduate-level students, statistics has become a requisite course for a wide range of fields. The learning of statistics cultivates students' quantitative and logical thinking, and additionally provides skills necessary for future employment choices. Proficiency in statistical skills has been observed to improve job performance because computational statistics tasks are routinely performed in most professions. As a result, statistics educators, researchers and specialists are motivated to form communities and committees focusing on developing curriculum guidelines for college-level statistics education. For example, in 1999, the Undergraduate Statistics Education Initiative (USEI), a committee of ten representatives from academia, industry, and the government, began meeting to suggest improvements for the college statistics curriculum (Bryce, 2005). Additionally, in 2003, the American Statistical Association published a curricular guideline for college-level statistics. In this curricular guideline, they suggested that the curriculum topics for undergraduate statistics minors consist of general statistical methodology (statistical thinking, descriptive, estimation, testing, and others), statistical modeling (simple and multiple regression, diagnostics, and others) and exposure to professional statistical software ("Curriculum guidelines for undergraduate programs in statistical science," 2010). This successful statistics curriculum is one of the most used references for college statistics instructors.

College statistics courses are expected to offer students tools for processing real world data: trimming the mass of information, ordering the messiness, separating sense from nonsense, and selecting the relevant few from the irrelevant many (Ben-Zvi, 2000). However, students face challenges in statistics courses. Researchers have found that many students do not obtain an adequate understanding of basic statistics concepts to solve applied problems (Garfield, 1995). Studies showed that students often respond to statistical problems by falling into a "number crunching" mode. That is, students plug quantities into a computational formula or procedure without forming an internal interpretation of the problem (Garfield & Ahlgren, 1988). Consequently, students may be able to memorize the formulas and the steps to follow, but seldom do they successfully develop the abilities of statistical reasoning and interpretation. Students should be able to render reasoned descriptions, inferences and opinions about data, or argue about the interpretation of data (Gal & Garfield, 1999). Without this ability, students cannot make appropriate decisions in applied real-world problems.

Researchers and educators have come up with multiple solutions to solve students' difficulties and enhance the learning outcomes in statistics. For example, a small-group cooperative environment was suggested for students to better construct knowledge (Garfield, 1995). The learning-by-doing and learning-by-practicing methods have also been popular in the statistics classroom (Smith, 1998). In the past ten years, with the advance of computer technology, the most recommended strategy to overcome students' difficulties in learning statistics is the integration of computer technology (Chance, Ben-Zvi, Garfield, & Medina, 2007). Technological tools are expected to support student learning by providing automation of calculations, visualization of abstract concepts, emphasis on data exploration, and investigation of real-life problems (Chance et al., 2007).

Among the possible technological tools, animated visual aids have high potential to promote learners' knowledge acquisition (Anglin, Vaez & Cunningham, 2004). Animated visual aids refer to a visual representation or a visual display in which the graphical elements being displayed change with time (Vekiri, 2002). Films, animation and simulation are all one type of the animated visual aids. The use of animated visual aids enables the demonstration of complex abstract ideas and provides multiple examples in seconds (Chance et al., 2007). With the help of animated visual aids, students are better able to explore and "see" statistical ideas, and teachers are better able to present abstract concepts to students.

The first study that examined the benefits of using animated visual aids in undergraduate statistics teaching was published in 1983 by Lock. The animated visual aids at that time were mostly text-based interfaces, including Minitab and graphing calculators. These interfaces allowed users to interact with the visualization or simulation, but the interaction approaches were very limited. In the 1990's, animated visuals with graphic-based interfaces appear. ELASTIC was one of the popular examples. Then in the 2000's, with the development of JAVA and Flash technology, the animated visuals are mostly internet-based interfaces. The Rice Virtual Lab in Statistics (http://onlinestatbook.com/stat_sim/sampling_dist/index.html), Intuitor (http://www.intuitor.com/statistics/), and ConStats (http://constats.atech.tufts.edu/) are exemplary websites offering free statistical visual aids with Internet-based interface. The functionality and interactivity of the animated visual aids are much improved now.

There are advantages of using animated visual aids in learning statistics: 1) Students are enabled to understand abstract concepts by viewing graphics, and 2) students are allowed to manipulate parameters and methods (Ben-Zvi, 2000). An animated visual aid can be powerful in conceptualizing abstract knowledge. However, not all animated

visual aids work equally well in classroom practice. Instructors are often unsuccessful in using them. In a recent meta-analytic research paper which examined 42 studies on the use of animated visuals, it was found that the use of animated visulas did not facilitate learning (Anglin et al., 2004). The failure of the animations in the classroom can be attributed to the following factors: 1) Lack of interactivity (or user-control functions) (Tversky, Morrison, & Betrancourt, 2002) and 2) mediation by individual differences such as prior-knowledge and attitude (Boucheix & Schneider, 2008; Hegarty & Just, 1989; Mayer & Gallini, 1990; Schnotz & Grzondziel, 1999).

Interactivity, or the user-control function, is important to achieve learner apprehension (Lowe, 2004). Interactive displays which provide the opportunity to stop, rewind, and restart. From the memory demand viewpoint, this creates less cognitive load for learners and should improve comprehension. With the interactive function, the animation program is no longer overwhelming. Though interactive programs that allow close-ups, zooming, and control of speed are more likely to facilitate perception and comprehension (Tversky, Morrison, & Betrancourt's., 2002), not all animation programs are created with appropriate interactivity. Also, due to technological limitations, studies of interactive animation did not commence until 2001. The research findings on animation interactivity are limited and more work needs to be done.

Individual differences, based on the literature, include prior knowledge and learning attitude (KOC, 2005; Mayer & Gallini, 1990). Taking prior knowledge as an example, some researchers have found that students with low prior knowledge benefit more from computer-based visualization programs than high-knowledge students (Mayer & Gallini, 1990). However, some researchers have found that high-knowledge users are more capable of locating relevant information in the computer-based program and can therefore take more advantage of it (Hegarty & Just, 1989). In Schnotz and Grzondziel's

study (1999), they found that the dynamic computer modules did not help students to reduce cognitive loads, but made the learning worse than for students who used the traditional static pictures. This result was explained by students' high cognitive ability and high prior knowledge which made the provision of animation useless. Though having different conclusions, these studies showed that prior knowledge does affects students' use of animation. Most of the research regarding individual differences addresses prior knowledge, however, there are several research studies that point towards the effect of attitude characteristics such as affect, passion, dislike, fear, satisfaction and frustration (KOC, 2005). Recognizing a more comprehensive set of learner attributes, such as those influenced by emotions and intentions, is useful in guiding the design of instructional solutions and environments that enhance the overall learning experience.

The above sections point out the importance of interactivity and individual differences on students' use of animated visuals. Previous studies leave many unanswered questions, such as: How do we design an animation program with appropriate interactivity? How do we create an interactive mechanism in the animation program? How do student individual differences impact their use of the interactive module? How do students' use of the interactive program impact their test scores? The answers to these questions are still inconclusive. The purpose of this study is to answer the above questions not just by examining whether the animation program helps students learn, but by also exploring how individual differences affect learners' use of interactive animation. This research hoped to increase understanding about the use of interactive animation in classrooms, especially in the statistics education area.

STATEMENT OF PROBLEM

The problem which this study addresses is: How does the animation interactivity affect novices' learning in introductory statistics? The animation program developed for

this study corresponds to the unit on *the principles of hypothesis testing*, which is one of the most difficult units for students (Smith, 2003).

RESEARCH QUESTIONS

Based on the above problem statement, this study concerns the effect of animation interactivity on students' cognitive function, confidence, and perception. The following are three sub-questions for this study:

1. Is there any difference in achievement improvement among students who use different interactive levels of an animation program?
2. Is there any difference in confidence improvement among students who use different interactive levels of an animation program?
3. Is there any difference in program perception among students who use different interactive levels of an animation program?

A laboratory experiment was conducted to answer the above questions. This lab experiment had 123 participants, and the grouping procedure was totally random. The treatment was rigorously designed and the experimental procedure was be fully controlled by the researcher.

SIGNIFICANCE OF THE STUDY

Interactive animation has become increasingly popular in computer-based instruction. Although research has been conducted on this particular topic for the past ten years, there are many problems left unsolved. Researchers can not yet make a conclusion on how interactive animation affects learning and why it has a different impact on different individuals. The need for advanced research on the effects of interactive animation is still pressing.

This study used Mayer's (2009) multimedia design principles as well as Kristof and Satran's (1995) interactivity theory as the theoretical underpinning. The study used this framework to explore how an interactive animation program helps learning by considering different levels of program interactivity. In addition, this study aimed to clarify the role of individual differences in instruction where the interactive animation is integrated. Also, it informed educational practitioners about how to optimize instruction with interactive animation by considering individual differences.

The interactive program under investigation was developed for a statistics course. Statistics is a required course in most college disciplines, but it was reported that students' understanding of statistical concepts was not as good as instructors expected (Garfield, 1995). With hopes of improving student learning on statistics, this study tended to inform statistics instructors or instructional designers how to guide student learning through the appropriate use of interactive animation. The animation program developed for this study corresponded to the unit on *Principles of Hypothesis Testing*, which is one of the most difficult units for students. The program developed in this study created an interactive environment for students to actively construct knowledge introduced in this particular unit. This study presented an innovative approach to teaching abstract statistical concepts, and it also aimed to explore how interactive animation influences student learning of introductory statistics. This research hoped to serve as a reference for future statistics curriculum designers, textbook publishers or instructors in curricula design and implement.

Chapter 2: Literature Review

Cognitive, constructivist and feedback theories provide different explanations for why interactive animation should be integrated into the classroom. From the cognitive perspective, interactive animation improves learning by helping information processing. From the constructivist viewpoint, interactive animation makes learning efficient by creating multiple representations of abstract knowledge for students to actively construct a mental model. From the feedback viewpoint, the interactive animation provides stimulus to impact learners' future behaviors. In addition, Mayer's (2009) multimedia learning theory and Kristof and Satran's theory of interactivity (1995) also contributed to this theoretical framework and served as guidelines for the creation of the interactive animation program in this study.

This chapter presents an overview and synthesis of these theories. The review begins with a discussion on the theoretical framework of multimedia learning. The second section discusses the instructional use of interactive visuals, including the use of computer-based animation. The third section discusses the use of interactive technologies in statistics education. The fourth section discusses educational models for cognitive evaluation. This review of literature presents the gap between the existing research and research questions in this study.

THEORETICAL FOUNDATIONS OF MULTIMEDIA LEARNING

The term *multimedia* is defined as presenting both words (such as spoken text or printed text) and pictures (such as illustrations, photos, animation, or video) (Mayer, 2009). Multimedia learning refers to learning from material presented in both the verbal and pictorial. *Animation* is an important component of multimedia (Mayer & Chandler, 2001). It brings three attributes to an instructional setting thus enhances learning

outcomes: 1) Visualization, 2) motion, and 3) trajectory (direction) (Rieber, 1990). Efficacy of animation depends on the learner's need for one or more of three attributes for successful completion of an instructional task. Being one component of multimedia, the use of animation in instruction is supported by cognitive theories and constructivist theories, including the information-processing theory, dual-coding theory, theory of abstract knowledge acquisition, and the feedback theory. The following section describes these theoretical foundations.

Information-processing theory

Information-processing theory is fundamental to multimedia learning. In the 1960's, cognitive psychologists began to study memory and proposed theories of information processing and knowledge representation. From the cognitive perspective, the most important question is "how the mind does what it does when it works." The information-processing model proposes that the brain contains certain structures that process information much like a computer does. For example, in learning psychology, "short term memory" and "long term memory," descriptors of how computers process information, refer to how human brains encode information (Atkinson & Shiffrin, 1968). Information-processing theory assumes that the human brain has three kinds of memory or stores: 1) sensory registers, 2) short-term memory, and 3) long-term memory. The sensory register is the short-term memory store for the senses such as touch, vision, or hearing. It is presumed to hold large quantities of information, but only for milliseconds. Short-term memory is also called working memory. This refers to a brain system that provides temporary storage and manipulates the information necessary for complex cognitive tasks. Long-term memory refers to a brain system that permanently stores, manages, and retrieves information for later use. Information stored as long-term memory may be available for a lifetime.

Relies on both long-term and short-term memory, learning occurs through a process. First the sensory register gets the information and holds it for a very short time. The register can hold about seven items. If the learner pays attention to a piece of information, it will enter the short-term memory. Then, the information stays in the short-term memory for 5 to 20 seconds. If the learner practices or processes the information, the information will be transferred to the long-term memory for storage and retrieval. Otherwise, the information is lost. Theorists also believe that the new information needs to be integrated into existing cognitive structures, called "schema," in the long term memory. Schema can be combined, extended or altered to accommodate new information.

Cognitive theorists emphasized the thought process behind behavior. Changes in behavior are observed, and they are used as indicators as to what is happening inside the learner's mind. Based on information-processing theory, some processing aids such as advanced organizers or instructional-based aids (e.g., educational multimedia) are helpful since those aids facilitate learners to integrate mental schema more efficiently.

Dual-coding theory

In late 1980's, dual-coding theory was developed based on information-processing theory. According to dual-coding theory, providing materials in both visual and verbal format enhances learning because it helps information processing (Paivio, 1991). Dual-coding theory proposes that there are two distinct but interconnected cognitive channels for processing and storing information in the human mind: The imagery system and the verbal system. They are structurally distinct because they store information in representational units that are modality specific: logogens and imagens. Logogens are organized successively in larger units. Therefore, they are subject to the constraints of sequential processing, which limits the amount of information that can be

processed at a time. Imagens are organized in a synchronous manner, which allows many parts of a mental image to be available for simultaneous processing. Although the two cognitive systems are functionally distinct, they are interconnected. Associative connections can form between verbal and visual representations, enabling the transformation of one type of information into the other. For example, people can associate the word *book* with a picture of a book and thus hearing the word may elicit a mental image (Vekiri, 2002).

Paivio proposed that mental representations have both visual and verbal modes and these two modes connect with each other. Mayer and Anderson (1992) then provided empirical evidences to support Paivio's dual-coding theory. They divided college students into four groups to view animated visuals of how a bicycle tire pump works. The groups were a words-with-pictures group, a words-only group, a pictures-only group, and a control group. The experiment was done to understand the effects of words and pictures on learning. The results showed that the words-with-pictures group outperformed the other groups on the overall recall test and the problem-solving test. The study showed that words and pictures should be linked together for effective learning.

Dual-coding theory now has been widely used in multimedia instructional design. By utilizing dual-coding theory and presenting both visual and verbal systems to learners, we assume that the presentation of information is optimized.

Abstract knowledge acquisition

Constructivism also contributes to this theoretical framework for interactive animation use. Constructivist theorist Spiro divided knowledge into two different types: Well-structured and ill-structured (Spiro & Jehng, 1990). He then offered theories of learning that emphasized the need to treat ill-structured knowledge domains differently from well-structured domains. Well-structured knowledge items and ill-structured

knowledge items exist in every subject. Instructors must distinguish between them in instructional design to enhance better knowledge acquisition for their students. The differences between well-structured knowledge and ill-structured knowledge are discussed below.

Well-structured knowledge refers to simple and fundamental knowledge with clear rules or principles. Strategies to deliver well-structured knowledge include textbooks, lectures or tutorials. Taking statistics as an example, the calculation of the mean, standard deviation and t-value could be considered well-structured knowledge since they have very clear procedures and rules.

Ill-structured knowledge, or abstract knowledge, refers to complex knowledge which has no regular principles. Ill-structured knowledge items have the following two characteristics: 1) Knowledge application requires the interaction among multiple mental models, and 2) knowledge item has cross-case differences. Strategies to deliver ill-structured items include problem-based learning, self-learning or discovery-learning. Again taking statistics for an example, the interpretation or evaluation of an experimental design is an ill-structured knowledge item since there are many factors related to the experimental design, including sample size, effect size, alpha level and practical significance. Sometimes "experiences" help learners to interpret such statistical data instead of rules.

It is comparatively difficult for teachers to deliver ill-structured knowledge since it has higher complexity and flexibility. Many educational researchers have conducted studies looking for a good strategy to make learning ill-structured knowledge efficient. Currently, Spiro's "cognitive flexibility theory" is dominant. Cognitive flexibility theory suggests the following solutions for teachers to better deal with ill-structured knowledge items: 1) Multiple representations, 2) revisiting the materials, 3) schema assembly, and 4)

management of complexity. The integration of technology is also mentioned as a solution to deliver ill-structured knowledge, and the computer ideally provides functions that foster cognitive flexibility (Spiro, Feltovich, Jacobson, & Coulson, 1992). In particular, computers' multidimensional and nonlinear presentations have the power to convey ill-structured aspects of knowledge domains in ways that traditional learning environments (textbooks, lectures, computer-based drills) cannot. In other words, multimedia or animated visuals will help students learn abstract knowledge since they provide multiple representations as well as a non-linear environment for students to construct knowledge.

Feedback theory

Feedback has been widely perceived as an essential component of general systems operations, including multimedia learning systems and animation systems. In the current educational setting, feedback can be said to describe any communication or procedure given to inform a learner of the accuracy of a response, usually to an instructional question (Mory, 2003). Behaviorist E. L. Thorndike's Law of Effect, developed in one of the earliest feedback studies, mentioned that feedback can act as a "connector" between stimuli and responses (Kulhavy & Wanger, 1993). With the advance of technology, new learning environments offer a wide range of potential feedback uses that were not previously considered. For example, technological tools such as simulation, animation or games could be used to provide learners with feedback.

From a behaviorist viewpoint, feedback uses the discriminative stimulus to reinforce the consequences of behavior and impact future behavior. This principle is grounded in the design of programmed instruction, which prompts the learner to emit a target behavior and then reinforces it (Bangert-Drowns, Kulik, Kulik, & Morgan, 1991). Cognitive theorists also emphasized the consequences of behavior, but they saw feedback as more than a mechanism to strengthen automatic stimulus-response association.

Feedback was viewed as necessary for verification of retrieval accuracy, concept development, skill refinement, and meta-cognitive adaptation.

Several areas are well represented in the research literature on feedback, such as 1) complexity of feedback, 2) timing of feedback (immediate versus delayed), and 3) using feedback for error analyses (Mory, 2003). Feedback complexity refers to how much and what information should be included in the feedback messages. Dempsey, Driscoll and Swindell (1993) have organized feedback complexity as follows: 1) No feedback, 2) simple verification feedback or knowledge of results, 3) correct response feedback or knowledge of correct response, 4) elaborated feedback, and 5) try-again feedback. Kulhavy and Stock (1989) have organized feedback from an elaboration perspective: 1) Task-specific feedback, 2) instruction based feedback, and 3) extra-instructional feedback. Although there is an abundance of literature concerning feedback complexity, the research is inconclusive as to which type of feedback is most helpful to learners. Only half of the studies using task-specific feedback produced significant improvements in learning (Mory, 2003). Inconsistency has also been found in studies using information-based feedback. Such variance has made it hard to prescribe any set rules for the feedback use.

The timing of feedback is another issue on which researchers have focused. Researchers have organized feedback into two types based on timing: 1) Immediate feedback, and 2) delayed feedback (Mory, 2003). These two types of feedback are frequently used in the design of computer-based instruction. Dempsey and Wager (1988) defined the two types of feedback as follows: immediate feedback gives informative corrective feedback to a learner during instruction as quickly as the computer's hardware and software will allow; delayed feedback gives informative, corrective feedback to a learner during instruction after a specified programming delay interval. Researchers have

concluded that in most learning situation, delayed feedback hinders the acquisition of needed information. Only in very special experimental situations has the use of delayed feedback helped learning (Mory, 2003).

Researchers are also interested in the error analyses which feedback can provide. Errors are viewed as valuable opportunities to clarify learner misunderstanding (Mory. 2003). As a result, errors play an important role in feedback studies today. Phye and Bender (1989) have suggested that feedback has three types of functions from the error analysis perspective: 1) Confirmatory function, 2) corrective function, and 3) no function. The confirmatory function means that feedback serves to confirm a correct answer to a pre-test question, and the corrective function serves to correct an error made on a pre-test question. If errors result on the post-test questions, feedback can be described as no function. Phye and Bender have suggested that designers be careful and clear about the function which feedback provides.

Feedback complexity, feedback timing, and error analyses play important roles in all aspects of instructional system design. Based on the literature, appropriate feedback makes the whole learning process more student-controlled, timely, and supportive. Also, learners are expected have more positive outcomes with the provision of feedback with high interactive levels (Mory, 2003). These feedback theories serve as a fundamental framework for this study in the creation of an interactive mechanism with different levels of feedback in the interactive animation.

INSTRUCTIONAL USE OF INTERACTIVE VISUALS

Computer technology is now more available and affordable than ever before. Many educational materials are now provided in commonplace technologies, such as videotape, CD, DVD, and World Wide Web. However, the new forms of multimedia technology will not educate students without high-quality content. To help students learn

difficult concepts, multimedia needs interactive capabilities for simulation and visualization. Such interactive, dynamic representations are the core content of educational learning modules.

Computer-based visual aids in education

The computer-based visual aid is one of the most powerful interactive multimedia tools for learning. The term *visual aid* refers to a visual representation or a visual display facilitating student learning. Visual representations and visual displays are used interchangeably in the literature to characterize displays that represent objects, concepts, and their relations using symbols and spatial arrangements (Vekiri, 2002). Visual aids are frequently integrated into the classroom setting to support learning and to make the learning process more effective. The purpose is to provide the viewer with visual means of processing information (Segenchuk, 1997).

Since 1980, the visual aid has been viewed as an innovative strategy in education, especially for math or science related subjects. Different types of visual aids have been developed to enhance students' concept formation. In computer science, for example, the impetus for visual aids comes from the abstractness of the concepts in the field. To make these concepts more concrete, graphical representations have been suggested to help one to better understand how concepts work (Naps, Rößling, Almstrum, Dann, Fleischer, Hundhausen, Korhonen, Malmi, McNally, Rodger, & Velazquez-Iturbide, 2003). Also, in physics and chemistry classes, it is very common for teachers to integrate visualizations to show molecular structure. To give students a clear understanding of molecular modeling, instructors can show a molecular simulation or engage students in creating a molecular simulation in the class (Allen, 2007). In a biology class, the instructor can use a simulated frog dissection as preparation for an actual dissection or as a substitute for the dissection (Kulik, 2002). Medical science is also a field which relies on visual aids;

simulation, one type of visual aid, is used to train medical practitioners to diagnose patients (Savoldelli et al., 2005). The above examples show the possibilities for using visual aids in educational settings.

There are several ways to categorize and organize the terms related to visual displays. One of the most popular approaches is to categorize visual displays into two main systems: pictorial representations and graphical representations (Tversky et al., 2002). Pictorial representations are those in which interpretation involves subjectivity and ambiguity, such as paintings, photographs, and drawings. Conversely, graphical representations have a one-to-one correspondence between their elements and their referents, and each element has only one meaning (Vekiri, 2002). Some examples of graphical representations include diagrams, maps, graphs, charts or simulation. Another approach to categorization is to divide visual aids into two groups: static visuals and animated visuals (Anglin et al., 2004). The following section discusses the differences between them in detail.

Static visuals

In static visuals, the graphical elements being displayed can not change with time (Vekiri, 2002). Pictures, photos or graphs are all examples of static visuals. A sample of the static visual is the Digital Anatomist Project created by the University of Washington, which included digital atlases of human organs to facilitate student learning (Structural Informatics Group, 2007). According to Segenchuk (1997), static visuals are capable of drawing an analogy to another domain of knowledge, serving as an intermediate representation to bridge the learners' prior knowledge and new knowledge. These visuals make the important relationships within a concept salient.

Researchers began to explore the educational value of static visuals in the 1930's. Many functional frameworks were presented for categorizing different types of static

visuals. Duchastel (1978) identified three general functional roles of static visuals in texts: 1) An attentional role, 2) a retentional role, and 3) an explicative role. The attentional role relies on the fact that pictures naturally attract attention; the retentional role aids the learner in recalling information seen in a visual aid; the explicative role explains information that would be difficult to convey in verbal or written terms. Levie and Lentz (1982) provided a four-major-function framework for static visuals: 1) Attentional, 2) affective, 3) cognitive, and 4) compensatory. The attentional function refers to the visual aid attracting or directing learners' attention to the material; the affective function refers to the visual aid enhancing learners' enjoyment; the cognitive function refers to the visual aid facilitating content learning through improving comprehension, retention and additional information; the compensatory function refers to the visual aid accommodating struggling readers. Levin (1981) suggested a five-function framework: 1) Decoration, 2) representation, 3) organization, 4) interpretative, and 5) transformation. The decoration function refers to the association of texts with text-irrelevant pictures to attract attention; the representational function refers to the association of text-relevant pictures for better representation; the organizational functional refers to the provision of pictures as an organizational structure giving the texts more coherence; the interpretational function refers to the provision of pictures as capable of clarifying concepts or ideas for content that is hard to understand; the transformational function refers to the provision of pictures directly impacting learners' memories. Alesandrini's (1984) framework of static visuals had three functions: 1) Representational, 2) analogical, and 3) arbitrary. Representational pictures can convey information in a direct way through tangible objects or concepts (e.g., photos or drawings); analogical pictures convey meaning by acting as a substitute and then implying a similarity to the concept or topic being presented (e.g., logical pictures);

arbitrary pictures are highly schematized visuals that do not look like the things they represent but are related in some conceptual or logical ways (e.g., flowcharts, tree diagrams, maps and networks).

The above functional frameworks clearly define the instructional roles and benefits that static visual aids can provide. However, the static visual aid is still somewhat limited by its static property since it can not change with time. In recent years, more researchers and educators have turned to the animated visual aid to enhance student learning. The animated visual aid is viewed as a more dynamic and interactive learning tool for students. The following section discusses the properties of animated visuals.

Animated visuals

Animated visuals are those in which the graphical elements being displayed can change with time (Vekiri, 2002). Film, animation and dynamic visualization are all examples of animated visuals. The study of animated visual aids began in the 1920's with the instructional film. With the progress of technology development, animated visual aids expanded from film to computer-based or internet-based software programs. Use of this type of visual aid is increasing in educational materials across a range of subject disciplines and levels of study. The two main examples for animated visuals are animations and simulations.

Animated visuals bring three attributes to an instructional setting: Visualization, motion, and trajectory (trajectory refers to an animated object's path of travel) (Rieber, 1990). According to Rieber, the efficacy of an animated visual depends upon the learner's need for one or more of these attributes to successfully complete an instructional task. If only adequate visualization of an instructional task is needed, either a static or an animated visual will be sufficient, and no differentiation in effect should be expected.

In terms of the functional framework of animated visuals, Rieber (1990) suggested that "generally, animation has been used in instruction to fulfill one of three functions: attention-gaining, presentation and practice." More recently, Park and Hopkins (1993) have identified five important instructional roles for animated visuals: 1) As an attention guide 2) as an aid for illustration, 3) as a representation of domain knowledge, 4) as a device model for forming a mental image, and 5) as a visual analogy or reasoning anchor for understanding abstract and symbolic concepts or processes.

Both static visuals and animated visuals are capable of facilitating student learning of abstract concepts. However, the animated visual aid is viewed as more powerful than the static one because it engages learning in much more interactive ways.

Before computer technology had matured, it was very time and effort consuming for researchers or educators to develop animated visuals for classroom or research use. With the progress of technology such as Flash and JAVA, however, the creation of animated visuals is comparatively easier. This simplification has increased the number of studies conducted on animated visuals.

Interactivity and learner individual difference

The use of interactive animation can enhance student comprehension, learning, memory, communication and inference. However, research showed that only carefully designed programs are beneficial for conveying complex concepts. Researchers have suggested that instructional designers should adhere to principles to optimize visual design. Mayer (2009) derived ten design principles for multimedia learning by conducting a series of empirical studies based on the dual-coding theory. These ten principles provide practical guidelines for instructors or multimedia producers to develop good instructional multimedia.

1. Coherence principle: Students learn better when extraneous words, sounds, and pictures are excluded.
2. Signaling principle: Students learn better when cues that highlight the organization of the essential material are added.
3. Redundancy principle: Students learn better when words are presented as narration rather than as narration and on-screen text.
4. Spatial contiguity principle: Students learn better when corresponding words and pictures are near rather than far from each other.
5. Temporal contiguity principle: Students learn better when corresponding words and pictures are presented simultaneously rather than successively.
6. Segmenting principle: Students learn better when a multimedia message is presented in user-paced segments rather than as a continuous unit.
7. Pre-training principle: Students learn more deeply from a multimedia message when they know the names and characteristics of the main concepts.
8. Modality principle: Students learn more deeply from pictures and spoken words than from pictures and printed words.
9. Multimedia principle: Students learn better from words and pictures than from words alone.
10. Personalization principles: Students learn better from multimedia presentations when words are in conversational styles rather than formal styles.

Mayer's experiments and arguments mostly focused on the effect of static media (e.g., static diagrams and printed texts). Interactive multimedia (e.g., interactive visualization or animation) was involved in some of the studies, but was not discussed at length. The design principles for interactive media are different from Mayer's general

design principles for static media. Two principles for animated designs are presented in the next section: interactivity principle and individual difference principle.

Interactivity principle

Interactivity is an important benefit of multimedia Web-based learning instruction. Interactive displays provide the opportunity to stop, rewind, and restart. From the memory demand viewpoint, this produces less cognitive load in learners and should improve comprehension. Draves (2000) suggested that the quality of interactivity is more important than content for successful learning; Sim (1997) believed that interactivity plays a crucial role in knowledge acquisition and the development of cognitive skills; Tversky et al. (2002) have argued that interactivity can help overcome the difficulties of perception and comprehension during the learning process. Also, Kristof & Satran (1995) mentioned that interactivity motivates users by providing clear guidance and options, giving users controls that allow them to self-direct, and making the experience easy and intuitive. It is widely believed that interactivity is one of the most important and fundamental factors that affects student learning and attitudes in multimedia Web-based learning environments (Moore, 1989).

Definition of interactivity

Interactivity is defined as the degree to which participants in a communication process have control over, and can exchange roles in mutual discourses (Williams, Rice, & Rogers, 1989). In the education field, the terms interaction and interactivity are used interchangeably to refer to the communication between 1) student and subject content, 2) student and instructor, or 3) student and student (Moore, 1989). Though each type of interaction has different impacts on student learning, the student-content interaction has been argued as the most important interaction among all in the multimedia Web-based learning environment (Gao & Lehman, 2003).

The student-content interaction can be further categorized into student-initiated interaction and computer-initiated interaction. In the former, the student seeks some information from the content in a way similar to looking something up in a book. In the latter, the system requires some input from the learner. The interactions which exist in the computer-based multimedia are mostly computer-initiated interactions. Evans and Sabry (2002) have presented a three-way model to illustrate the computer-initiated interactivity in the multimedia environments. According to the three-way model of computer-initiated interactions, an interaction involves a sequence of three actions: initiation, response and feedback. Each action involves an exchange of information between the content (computer system) and the learner.

1. Computer initiation: The computer system invites inputs from the learner.
2. Learner response: The learner provides inputs.
3. Computer feedback: The computer system passes back information as a consequence of the response.

For example, a navigation interaction has these three actions: 1) The computer system presents buttons or a control to the learner, 2) the student presses the buttons or uses the control, and 3) the computer system presents a new screen to the learner. A question and answer interaction has three actions: 1) The computer system presents a question to the learner, 2) the student chooses from a number of options, and 3) the computer system tells the student why they are right or wrong.

The three actions of the three-way model form part of an interactive cycle. Now this model is used by trained experts to evaluate system interactivity (Evans & Sabry, 2002).

Level of interactivity

Interactivity ranges from low to high levels based on the quality of interaction. Low-level interactivity means less interaction, engagement, and cognitive processing. In contrast, high-level interactivity means more interaction, engagement, and cognitive processing. Researchers have suggested different ways to define the levels of multimedia interactivity. Kristof and Satran (1995) have divided the degree of interactivity into a seven-level scale based on the kind of control available to users (Table 2.1).

In addition, Schwier and Misanchuk (1993) have identified levels for interaction based on instructional quality. The levels are reactive interaction, proactive interaction, and mutual interaction.

1. Reactive interaction: A reactive interaction is a response to a presented stimuli, or an answer to a given question.
2. Proactive interaction: A proactive interaction emphasizes learner construction and generative activity.
3. Mutual interaction: A mutual interaction uses an artificial intelligence or virtual reality design, where the learner becomes a fully franchised citizen in the instructional environment.

The relationships among these levels are hierarchical in terms of the quality of the interaction. In other words, mutual interaction has a better interactive quality than reactive interaction because there is greater opportunity for mental engagement and learner investment (Schwier & Misanchuk, 1993).

Table 2.1 Continuum of interactivity (based on Kristof & Satran, 1995; Teo, Oh, Liu, & Wei, 2003)

Available control	Example of Interaction	
Control pace	Click to advance to the next item	Increasing level of interactivity as user control
Control sequence	Choose where to go at any time	
Control media	Start/Stop video; search text; scroll or zoom the view	
Control variables	Change the outcome of a chart; customize a database search	
Control transaction	Enter a password; send a message	
Control objects	Move items around screen	
Control simulation	Change the perspective or course of action	

Once the interactive levels were defined, researchers began to explore the effects of different interactive levels on student learning by providing empirical evidence. For example, Gao and Lehan (2003) have found that students who used instructional material designed with reactive and proactive interactions outperformed those who used traditional material without interactions on an achievement test. Moreover, Evans and Gibbon (2007) have found that students who used the interactive system (with control of pace, self-assessment questions and an interactive simulation) outperformed those who used a non-interactive system in a problem-solving test. They have also found students needed less time to complete both memory and problem-solving tests. Rieber (1990) found that students learning Newton's Laws of Physics scored higher on tests after using an interactive animation (an animation followed by multiple-choice questions including feedback). In addition, Kopcha and Sullivan (2008) have found that high-prior-

knowledge students achieved better when their preference for interactivity was matched with the type of program they used. There is one assumption underlying all of these interactivity theories: the higher the interactive level, the better the product. In other words, learning should increase when learners use interactive multimedia systems instead of non-interactive systems. Researchers have refered to this cognitive consequence as the *interactivity effect* (Evans & Gibbons, 2007).

Though interactivity is crucial in multimedia design, the empirical research findings showed contradictory results. Some studies have claimed that user-control functions are helpful to learning (Rieber, 1990) while others have not (Boucheix & Schneider, 2008). Boucheix and Schneider (2008) have presented substantial experimental evidence showing that providing user-control could not guarantee a positive learning result. The user might not be able to manipulate user-control functions due to a lack of prior knowledge or information management skills. In other words, interactivity design needs to match individual differences such as the user ability or skills.

Individual difference principle

Individual differences are reflected in how people think, feel, and behave (Somyurek, Guyer, & Atasoy, 2008). Educators have to understand and identify how individual differences influence learning to maximize the efficiency of multimedia. The individual differences which this study addresses include: 1) Prior knowledge, and 2) attitude characteristics.

There is a consensus among researchers that a learner's prior knowledge will influence his use of multimedia. However, research examining how learner characteristics influence the use of multimedia is still very contradictory. Some researchers have found that students with low prior knowledge benefited more from multimedia than high-knowledge students (Mayer & Gallini, 1990). In contrast, some

have found that high-knowledge users were more capable of locating the relevant information in the computer-based visual aid and therefore able to take better advantage of it (Hegarty & Just, 1989). Schnotz and Grzondziel's study (1999) found that multimedia did not help student to reduce cognitive load; the student learning was no better than students who used static pictures. They explained this result by arguing that students had enough cognitive ability or prior knowledge to deal with static pictures, therefore, the animated visuals were no longer necessary. In other words, student prior knowledge mediated the effect of the multimedia. The studies showed that, on the one hand, students needed to have a minimum of prior knowledge or some general relevant knowledge in order to interpret information provided in the computer-based visual aid. On the other hand, they might benefit more when their knowledge less advanced (Vekiri, 2002). It was possible that complex multimedia is effective for high-knowledge students whereas low-knowledge students benefit more from simpler multimedia with less information. From this perspective, the mismatching of learners' prior knowledge with the information provided by multimedia will negatively impact learning.

Although most of the research regarding individual differences addresses prior knowledge, there are several research studies that point towards the effect of attitude characteristics such as affect, passion, dislike, fear, satisfaction and frustration (KOC, 2005). Recognizing a more comprehensive set of learner attributes, such as those influenced by emotions and intentions, is useful in guiding the design of instructional solutions and environments that enhance the overall learning experience.

How learners' individual characteristics impact the effective use of interactive visuals needs to be considered (Anglin et al., 2004). Unlike research pertaining to static media, which encompass many additional studies and dozens of treatment conditions,

research on the effects of interactive visuals is relatively limited. More research needs to be completed concerning the functions of interactive visuals in learning materials.

LEARNING STATISTICS WITH INTERACTIVE TECHNOLOGY

The aim of this study was to use interactive animation to improve student learning of statistics. Therefore, the development of statistics education was a necessary part of the literature review. The following section starts from how students develop concepts and reasoning ability in statistics, and then discusses existing technological developments in statistical multimedia.

Concept development in statistics

Statistics has become requisite for a wide range of fields of study in the university, especially in the area of social sciences. It cultivates students' reasoning and problem-solving abilities such as comprehension, planning, execution, evaluation and interpretation (Garfield, 2002). Also, proficiency in statistical skills has been observed to improve job performance because the computational tasks are routinely performed in most professions. Due to the importance, statistics educators, researchers and specialists are motivated to develop curriculum guidelines and conduct related studies for college-level statistics education (Bryce, 2005; "Curriculum guidelines for undergraduate programs in statistical science," 2010; National council of teachers of Mathematics, 1989). The body of studies related to teaching and learning statistics is growing, however, many statistics teachers continue to express difficulty in helping students understand the course material (Garfield , 1995; Garfield & Ahlgren, 1988).

The introductory statistics course is usually divided into three areas for college-level students: descriptive statistics, probability theory, and inferential statistics (Garfield & Ahlgren, 1988). Over the past 20 years, the literature has been filled with comments by

instructors about students not attaining an adequate understanding of basic statistics concepts and not being able to solve applied statistical problems (Garfield & Ahlgren, 1988). Garfield (1995) described students' misunderstanding in terms of a variety of fallacies: gambler fallacy, base-rate fallacy, availability fallacy, conjunction fallacy and causality fallacy. The gambler fallacy is the tendency to view chance as a self-correcting process by using a representative heuristic. For example, after observing a long run of heads in coin tosses, most people believe that now a tail is 'due' because the occurrence of a tail will result in a more representative sequence than the occurrence of another head. The base-rate fallacy is the tendency to ignore base rates (prior rates) in favor of case-specific information, rather than an integration of two. The conjunction fallacy is peoples' tendency to judge two correlated events as occurring together (in "conjunction") rather than occurring alone. The causality fallacy is the tendency to think correlation implies causation.

In additional to above general fallacies, students have some specific learning difficulties to statistics. Falk (1986) mentioned that students often confuse the conditional probabilities P(R|H0) and P(H0|R) when they learn hypothesis testing, and students tend to fail to correctly interpret the meaning of α, β and the *p*-value. Most students are able to follow the basic procedures for hypothesis testing, however, even after discussing what Type I and II errors signify, students inevitably ask, "but where do you get alpha (α) (Smith, 2003)?" Table 2.2 shows a summary of the typical contents in college-level statistics course and students' misunderstanding in each area.

Prior research gave the following principles and suggestions to solve students' misunderstanding and difficulties in learning statistics (Garfield, 1995):

1. Students learn statistics by constructing knowledge.
2. Students learn statistics by active involvement in learning activities.

3. Students do well only when they practice what they are learning.
4. Statistics learning is enhanced by having students become aware of their misconceptions.
5. Calculators and computers should be used to help students visualize and explore data, not just to follow algorithms to predetermined ends.
6. Students learn statistics better if they receive consistent and helpful feedback on their performance.

Table 2.2 Students' misunderstanding in different statistics units

Area	Typical contents	Students' misunderstanding
Descriptive statistics	- Measures of central tendency - Measures of variability - Measures of position - Frequency distributions and graphs	
Probability theory	- Rules (addition, multiplication) - Independent and mutually exclusive events - Random variables - Probability distributions - The binomial distribution - The normal distribution - Sampling - Central limit theorem	- Gambler fallacy - Base-rate fallacy - Fallacy of the variability of sample means - *p*-value fallacy
Inferential statistics	- Estimating parameters - Testing hypotheses	- Fallacy of the transposed conditional - Conjunction fallacy - Casualty fallacy

The fifth principle directly relates to the use of computers. Many studies emphasized technological tools as a solution for enhancing students' statistical

understanding. Moore (1997) claimed that incorporating technology creates a visualized, cooperative, and activity-based environment; this aids students in learning high-level cognitive content. Chance et al. (2007) pointed out that technological tools, including statistical software packages, educational software, spreadsheets, applets, graphing calculators, multimedia materials and data repositories, can help students understand and interpret statistical ideas. These studies have stressed that interactive technology is potential to solve students' learning obstacles in statistics and to decrease the learning curve. Therefore, this study focused on this type of tool to explore how it could be used in statistics education. The following section will provide a historical review on how interactive technology is used in statistics education.

The use of interactive technology in statistics education

The 1980's : programs with text-based interface

Minitab In the 1980's, Minitab was one of the most popular tools for visualization and simulation in the statistics classroom. It was written specifically as an educational tool rather than for statisticians in industry or research, which educated for its popularity. Minitab allowed the users to compute, explore, and simulate (Hubbard, 1992). Teachers used it to simulate distribution (Dambolena, 1986; Dambolena, 1986), Bayesian method (Albert, 1993) and central limit theorem (Gordon, 1987).

Graphing calculator The graphing calculator was another popular tool in this period. Graphing calculators could support such topics as descriptive statistics, sampling, regression and correlation (Mittag & Taylor, 1996; Windsor, 1998). The research findings showed that because the graphing calculator allows students to manipulate real data, they improve student achievement and attitude (Mittag & Taylor, 1996).

In addition to the above tools, Basic package, Pascal package, Monte Carlo and Spreadsheet were also used as tools in statistics education. The commonality among the 1980's tools was the text-based interface design. Users were allowed to interact with the program. However, options for the interaction were mainly limited to keyboard-operations as opposed to drag-and-drop manipulation.

The 1990's : programs with graphic-based interface

ELASTIC Educational software with a graphic-based interface was developed in the 1990's due to advances in technology and programming languages. ELASTIC (Environment for Learning Abstract Statistical Thinking) was one of the most popular software programs; it used the interactive graphics to teach fundamental statistical concepts. ELASTIC was an easy-to-use tool for entering, manipulating and displaying data; users could easily resize and reposition any graph on the screen (Rosebery & Rubin, 1989, 1990). It also coupled the power of a database management system with innovative graphing capabilities. Students could choose from histograms, bar charts, box plots, and scatter plots to display data.

In addition, some researchers developed graphic-based interface software using languages such as LISP and C++, which worked similarly to ELASTIC (Groeneboom, Jong, Tischenko, & Zomeren, 1996 ; Marasinghe, Meeker, Cook & Shin, 1996). Users were allowed to interact with graphic-based interfaces by using the 1990's tools, however, these tools were not capable of providing networked or distributed learning activities.

The 2000's : programs with Internet-based interface

JAVA applications JAVA applets are computer applications designed for the Internet. Applets are platform-independent, meaning they can run on any operation

system that has a JAVA Virtual Machine to translate applet byte codes to the appropriate platform-dependent instructions. The advantages of JAVA applets include its better performance in speed, interoperability, user interaction, portability, and network computing (Kamthan, 1999). Some noted JAVA applets for statistics education are Rice Virtual Labs in statistics, Elementary Statiscs JAVA applets, Statlets, StatCrunch, Statiscope, PsychStat, BusinessStat, Probability by Surprise, Web Interface for Statistics Education, CUWU Stats, PSOL, StatLab, Virtual Labs in Probability & Statistics, JavaStat, Vestac, JSci and CyberStats (Dinov, Sanchez, & Christou, 2008).

Flash applications Flash applications are the latest way to create educational visual aids and simulations in statistics. ConStatS, developed by Tufts University, consists of twelve programs, divided into five distinct parts: representing data, probability, sampling, inference, and experiments. It has a "point and click" graphic interface, developed with Flash. Students use pull down menus to access datasets, exercises, experiments, and different topics.

As mentioned, technological progress greatly enhances the potential for interactive visual aid use in statistics classrooms. With the use of interactive visuals, students now have the following advantages in statistics learning:

1. Students can understand abstract concepts by viewing graphics. The interactive technology support enhances accessibility of many statistical concepts by transforming purely symbolic presentations into spatial-geometric ones (Ben-Zvi, 2000).
2. Students can manipulate parameters and methods. The greatest advantage of this approach is that students can have full control over population properties, and can see how these properties are reflected in statistics (Velleman & Moore, 1996). Students can conceptualize statistics by manipulating

parameters and methods. Take the "sampling distribution" unit as example. The interactive visual enhances understanding because it allows them to 1) change data on the graphical representation, 2) manipulate the shape of a distribution on its numerical summaries, and 3) change the sample size on the distribution of the mean (Ben-Zvi, 2000). As a result, students can explore the nature of the sampling distribution by enlarging the sample size or changing the shape of the population distribution. Through manipulating inputs and parameters, students have more opportunities to recognize the properties of the sampling distribution.

3. Students can explore "what-if" questions. These tools allow students to answer "what happens if this is repeated a large number of times" through direct observation. Take Rice Virtual Lab's sampling distribution simulation as an example. To help students develop understanding to counterintuitive knowledge, their program creates simulations with different parameters (such as sample size, number of repetitions). And students can describe and explain their observations for these repeated simulations (DeMas, Garfield, & Chance, 1999).
4. Statistical concepts are connected to real-world analogies. Ben-Zvi (2000) used the term "Microworlds" to describe statistical education software which demonstrates concepts and methods by using interactive experiments, exploratory visualizations, and simulations. Manipulation in microworld-type software undoubtedly connects abstract statistical concepts to the real world.

Though the interactive technology is powerful in conceptualizing abstract knowledge, it has potential problems. For example, animated visuals may require explicit instructions for use. Students may need a real-time display, explanation or guidance to

better understand how multimedia work. More work needs to do for optimize the use of computer technology in statistics education.

COGNITIVE EVALUATION MODELS

Bloom's revised cognitive taxonomy and Garfield's statistical reasoning model were used in this study to examine student learning improvement. The following sections discuss and compare these two models.

A general model: Bloom's revised cognitive taxonomy

Benjamin Bloom developed the taxonomy for cognitive objects in the 1950's. Now, it is one of the most popular models guiding researchers and educators in achievement and test design. Bloom's taxonomy is a multi-level model that classifies thinking according to six cognitive levels of complexity. The levels have often been depicted as a stairway, leading teachers to encourage learners to "climb to a higher level of thought." The six levels are knowledge, comprehension, application, analysis, synthesis, and evaluation. In 2001, Anderson and Krathwohl revised the taxonomy and changed the terms from noun to verb forms. That is, the new taxonomy focuses more on the "cognitive process" (Anderson & Krathwohl, 2001). The six cognitive levels are described as follows (Anderson & Krathwohl, 2001):

Level 1. Remembering The learner is able to recall, restate and remember learned information. Some tasks include recognizing, listing, describing, identifying, retrieving, naming and locating.

Level 2. Understanding The learner grasps the meaning of information by interpreting and translating what has been learned. Some tasks include interpreting, exemplifying, summarizing, inferring, paraphrasing, classifying, comparing, and explaining.

Level 3. Applying The learner makes use of information in a context different from the one in which it was learned. In other words, the learner uses strategies, concepts, principles and theories in new situations. Some tasks include implementing and executing, computing and demonstrating.

Level 4. Analyzing The learner breaks learned information into parts to best understand that information. Some tasks include organizing, structuring and integrating.

Level 5. Evaluating The learner makes decisions based on in-depth reflection, criticism and assessment. In other words, the learner needs to judge the value of ideas, materials, and methods by developing and applying standards and criteria. Some tasks include critiquing, judging, experimenting and monitoring.

Level 6. Creating The learner creates new ideas and information using what has been previously learned. In other words, the learner needs to put together ideas or elements to develop an original idea or engage in creative thinking. Some tasks include designing, constructing, planning, producing, and inventing.

This revised taxonomy is widely used to evaluate learner learning outcomes. It is a classification of thinking organized by level of complexity, giving teachers a simple structure for different questions and learners an opportunity to practice a range of thinking skills.

Bloom's revised taxonomy is used in many educational fields including statistics education. In statistics education, in additional to Bloom's taxonomy, there is another specific model which is also very popular for examining learner learning. That is Garfield's statistical reasoning model. Garfield's statistical reasoning model shares similar properties with Bloom's taxonomy. Both of them categorize learners' cognitive skills and thinking into levels from low to high. The following section discusses Garfield's model in detail.

A specific model in statistics learning: Garfield's reasoning model

Statistical reasoning is defined as the way people reason with statistical ideas and make sense of statistical information (Gal & Garfield, 1999). This involves making interpretations based on data sets, graphical representations and statistical summaries (Garfield, 2002). In Garfield's 2002 study, she presented a model of statistical reasoning for helping learners exceed performance expectations in a one-semester introductory statistics course. Garfield created the model after conducting interviews with college learners enrolled in introductory statistics. The model describes a process of step-wise cognitive integration of statistical concepts, consisting of five levels as follows.

Level 1. Idiosyncratic reasoning The learner knows some statistical words and symbols, uses them without fully understanding them, often incorrectly, and may scramble them with unrelated information. For example, the learner has learned the terms mean, median, and standard deviation as summary measures, but uses them incorrectly.

Level 2. Verbal reasoning The learner has a verbal understanding of some concepts, but cannot apply this to the actual behavior. In other words, he memorizes and uses concepts or definitions correctly, but fails to interpret them and cannot apply them to real-world questions.

Level 3. Transitional reasoning The learner is able to correctly identify one or two dimensions of a statistical process without fully integrating these dimensions. In other words, he can interpret concepts and apply some of them to real-world questions.

Level 4. Procedural reasoning The learner is able to correctly identify the dimensions of a statistical concept or process but does not fully integrate them or understand the process. In other words, he can solve real-world problems without correct interpretations.

Level 5. Integrated process reasoning The learner has a complete understanding of the statistical process, and coordinates the rules and behavior. In other words, he can solve real-world problems with correct interpretations.

Garfield's model of statistical reasoning is consistent with Bloom's revised taxonomy of cognitive development (Table 2.3).

Table 2.3 Mapping Garfield's reasoning model to Bloom's revised cognitive taxonomy

Garfield's reasoning model	Revised Bloom's taxonomy	Elaboration/Example
Idiosyncratic reasoning	--	--
Verbal reasoning	Remembering	Learners are able to recall the definition or meaning of statistical concepts.
Transitional reasoning	Understanding	Learners are able to explain, paraphrase or exemplify the statistical concepts.
Procedural reasoning	Applying	Learners are able to solve application problems.
Integrated process reasoning		Learners are able to solve application problems with correct interpretation.

Both Garfield and Bloom's revised models provided guidelines for statistics instructors or researchers to carefully examine student learning. However, the development and evaluation of students' reasoning ability was much more emphasized in statistics education. According to Garfield (2002), unless students' reasoning is carefully examined, especially in applied contexts, they may not advance past the early stages of

reasoning and may not have the integrated understanding needed to make correct judgments and interpretations. Garfield offered suggestions for classroom assessment techniques to evaluate students' statistical reasoning, including case studies or authentic tasks, concept maps, critiques of statistical ideas or issues in the news, minute papers and enhanced multiple-choice items.

This study took advantage of both above models to examine student achievement improvement. The later chapter discusses the measurement for these cognitive levels of learning in detail.

SUMMARY

This study employed cognitive, constructivist, and feedback theories as a fundamental framework to explore the integration of interactive animation into the classroom. Cognitive theorists argued that the provision of animated visual tools are helpful for learning since it facilitate learners' information processing (Mayer, 2009; Rieber, 1990, 1991; Vekiri, 2002); constructivists suggested that providing learners with tools for multiple representations makes learning ill-structured or abstract knowledge efficient (Spiro et al., 1990); feedback theorists claimed that the interactive animation provides stimulus to impact learners' future behaviors (Mory, 2003). Educators have high expectations for the effect of interactive animation. However, the research findings are limited, especially in the field of statistics education. Some unsolved questions include: 1) How the use of an interactive animation program impacts student learning in statistics, and 2) how individual differences impact learning with an interactive animation program.

This study was designed to answer the above questions. In this chapter, design principles such as Mayer's (2009) multimedia learning design principles, and Kristof and

Satran's (2002) theory of interactivity are discussed. Chapter 3 describes the interactive animation program created for this study following the design principles in Chapter 2.

Chapter 3: Methodology

This chapter presents the animation program and research method used in the study. The first section describes the program design and development, including how the interactive animation program was created and why; the program interface is shown as well. The second section describes the research design and procedures, including the research hypotheses, pilot study results, subjects, treatment, experimental procedure, instruments, data collection and analysis.

DESIGN AND DEVELOPMENT OF AN INTERACTIVE ANIMATION PROGRAM

Learning objectives

An interactive animation program was created to facilitate learning *Principles of Hypothesis Testing*. The design was based on Dr. Brandon Vaughn's introductory statistics course for undergraduate classes at the University of Texas at Austin. In this unit, students were expected to distinguish the null hypothesis from the alternative hypothesis and apply type I error (α), type II error (β) and the *p*-value to make decisions in significance testing. The specific learning objectives included:

1. The student will be able to set up the null hypothesis as well as the alternative hypothesis for real-world problems.
2. The student will be able to understand the errors (type I and type II) which can occur in the testing process.
3. The student will be able to calculate the type I error (α), type II error (β), and *p*-value in real-world problems with different population shapes.
4. The student will be able to make a decision about a real-world problem based on the number of the type I error (α), type II error (β), and *p*-value.

In the traditional classroom, the instructor taught the above concepts using the textbook and blackboard. The following figure is a sample textbook page from this unit.

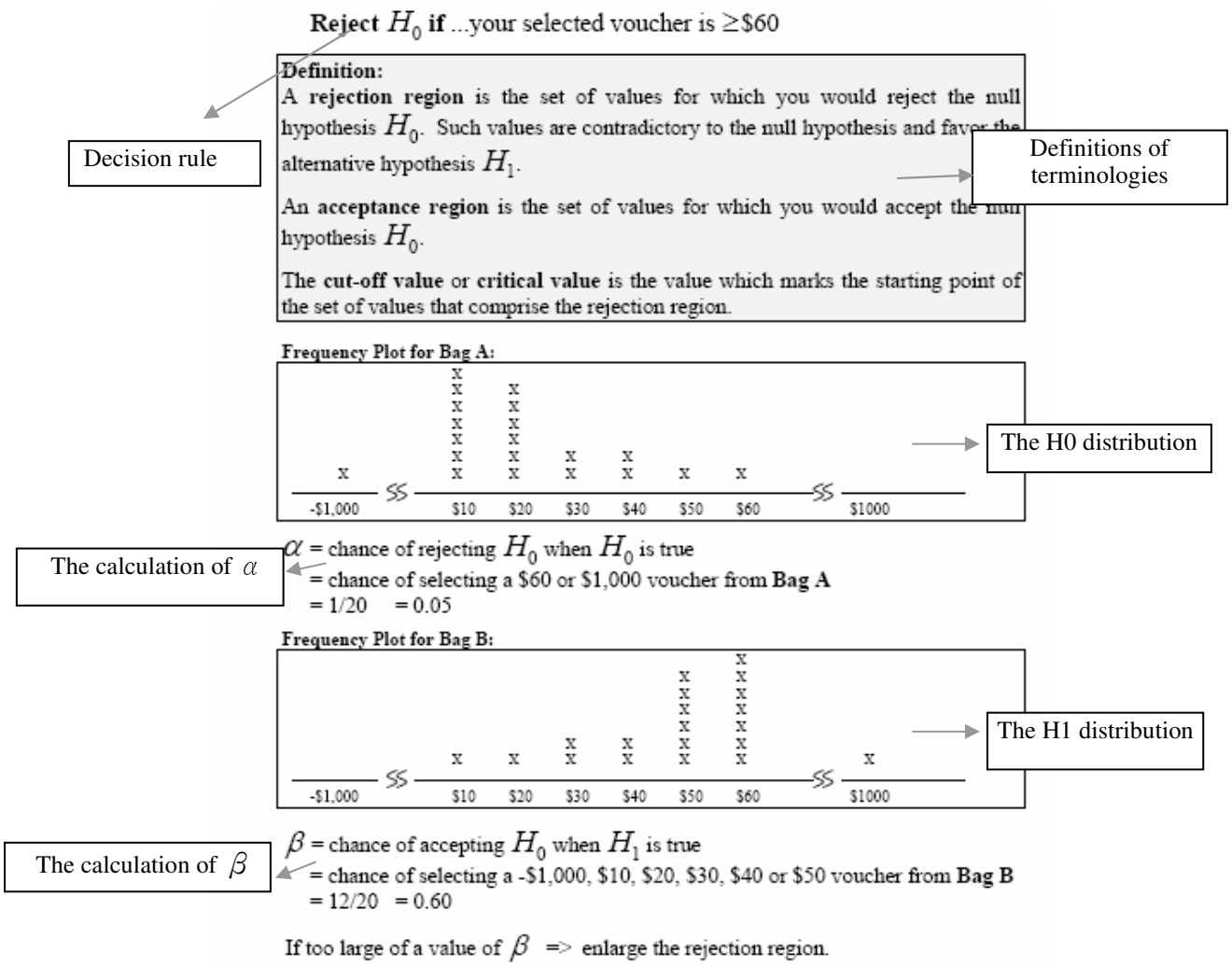

Figure 3.1 Sample page of a traditional textbook for statistical language of hypothesis testing

When students used textbooks, however, there was no way for them to interact with the numbers; students could only passively read the definitions and examples. Based on the premise that interactivity can enhance learning, student understanding of statistical

hypothesis testing might be limited if they were only passively engaged. The design of the animation program was based on Dr. Vaughn's introductory statistics textbook. Moreover, the program was created with interactivity therefore the users had the opportunity to interactive with the sample size, distribution type, and the decision rule for hypothesis testing. This program could be used as a supportive module in classroom teaching or used as a stand-alone self-learning module. In this study, this program was used as a stand-alone self-learning module.

Program design and development

Program interface

Adobe Flash CS3 Professional and Action Script 2.0 were used to create the interactive animation program designed as a 480px * 560px window. The small screen size made it possible for the instructor and students to install the interactive animation program on any mobile device; in this particular study, however, the laptop was the primary device used.

A user-controlled mechanism was built inside the program to fulfill the interactivity principle allowing users to experience different levels of interaction based on their needs and preferences (Tversky et al., 2002; Kristof & Satran, 1995). In addition to the Question Statement Page (Figure 3.2), users could interact with the program in four different modes: 1) Static Instruction, 2) Simple Animation, 3) Change Inputs, and 4) Practice. The four modes were selected mainly based on Kristof and Satran's continuum of interactivity (1995). The Animation Mode allowed learners to control animation pace and sequence, corresponding to level one and two on the continuum. The Change Mode allowed learners to control program variables, corresponding to level three on the continuum. The Practice Mode did not correspond to any level on Kristof and Satran's

continuum of interactivity; however, it was designed as a mode with the highest interactivity since learners could get feedbacks from the program. In addition to dynamic visuals for concepts of type I error (α), type II error (β) and the p-value, a female carton character was put on the screen as a virtual learning tutor. This cartoon character showed up on the Question Statement Page and Static Instruction Mode to give students hints for following tasks. The following sections describe these modes in detail.

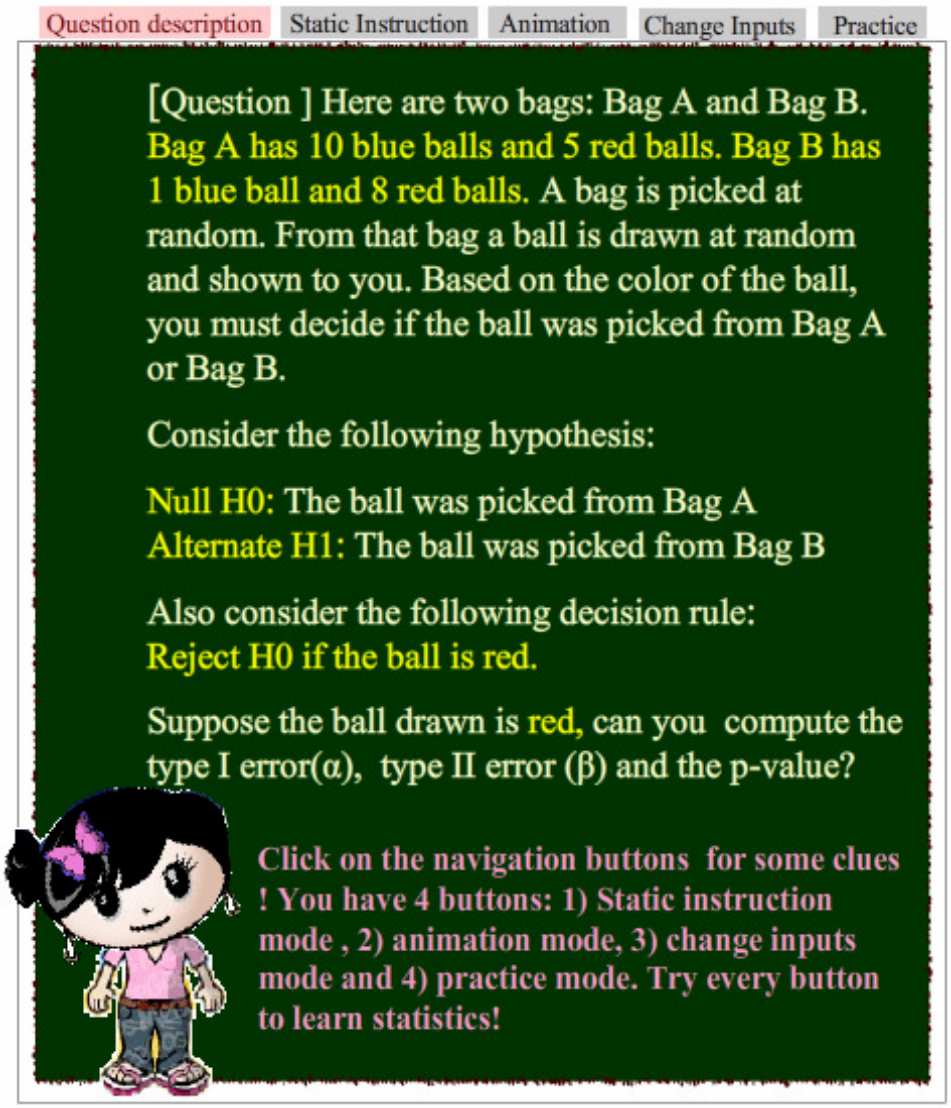

Figure 3.2 Program interface for the Question Statement Page

Static Instruction

The Static Instruction Mode provided the lowest level of interaction between the user and the program. The textual definitions and the computing procedures for the type I error (α), type II error (β) and *p*-value were put on the program screen as a static pdf file (Figure3.3). The presentation of this mode was similar to a traditional lecture.

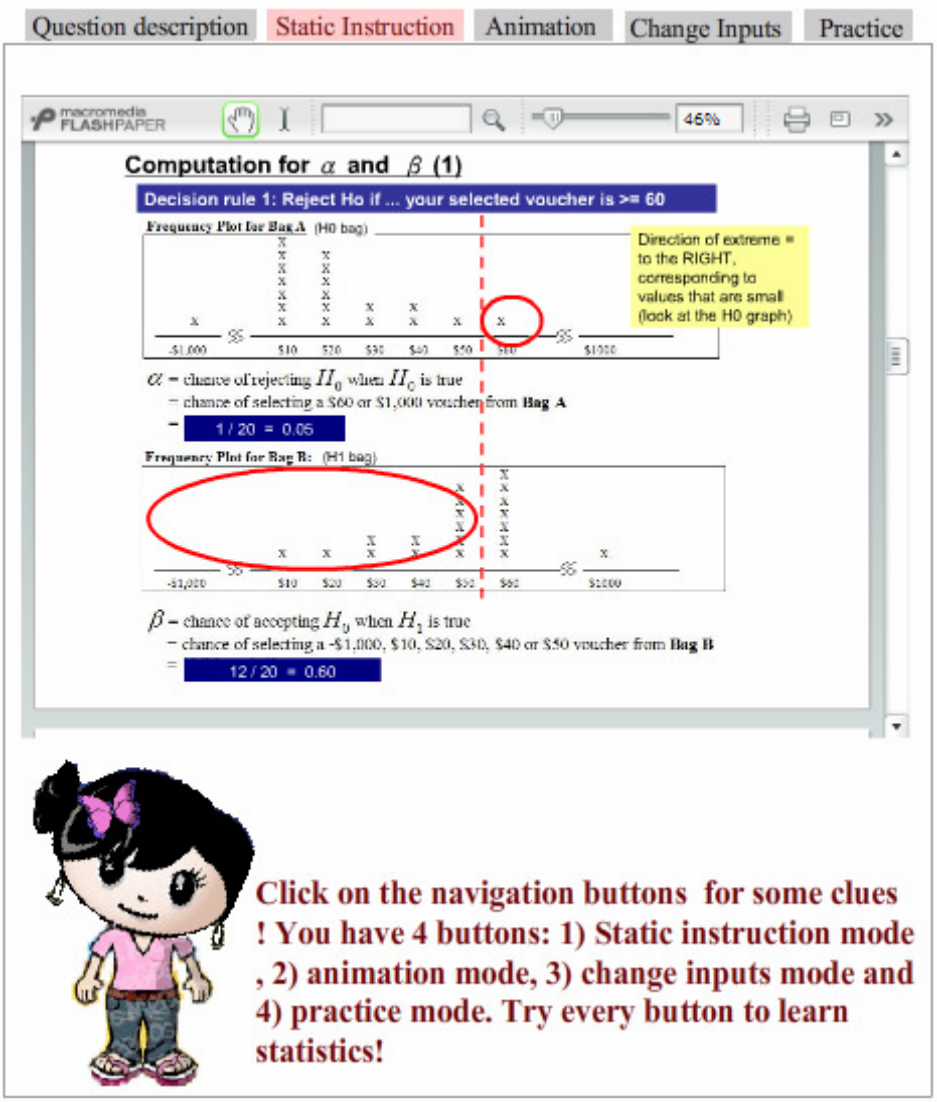

Figure 3.3 Program interface for the Static Instruction Mode

Simple Animation

The Simple Animation Mode allowed learners to control the sequence and pace of three types of animation (left-tailed, right-tailed, and two-tailed). Learners could repeatedly watch the animated procedures for calculating α, β, and *p*-values with fixed parameters. The program first displayed a question statement on the left side of the screen, and then the animated calculation to this particular question followed on the right side of screen. After the learner finished watching the animation, he was free to switch to new animations for other questions. Figure 3.4 shows the Simple Animation Mode interface.

Figure 3.4 shows the calculation procedure of a one-sided test. The program posits a question with some particular hypothesis settings and decision rule on the left side of the window. In this case, the H0 distribution is set up as a "left-tailed" distribution, the voucher drawn is set up as "30," and the decision rule is set up as "Reject H0 if value <= 40." The right-side window shows the animation of the calculation for α, β, and the *p*-value. Three graphs are used to illustrate the calculation. The top graph shows the calculating procedure for α. The black boxes represent the voucher values in the H0 bag; the red boxes represent those vouchers which fall into the rejection region. In this case, while calculating α, there are 20 boxes (vouchers) in the bag and only 6 fall in the rejection region. As a result, α equals six out of twenty (6/20). Similar calculation processes are animated on the middle and bottom graphs for β and *p*-value. After the calculations are finished, the testing conclusions are shown as texts in the bottom of the window. In this case, the program tells the learners "since *p*-value$<\alpha$, we reject H0." Under this mode, learners cannot change any of the parameters; they can only repeatedly watch the same question for three different types of tests (right-tailed, left-tailed, and two-tailed).

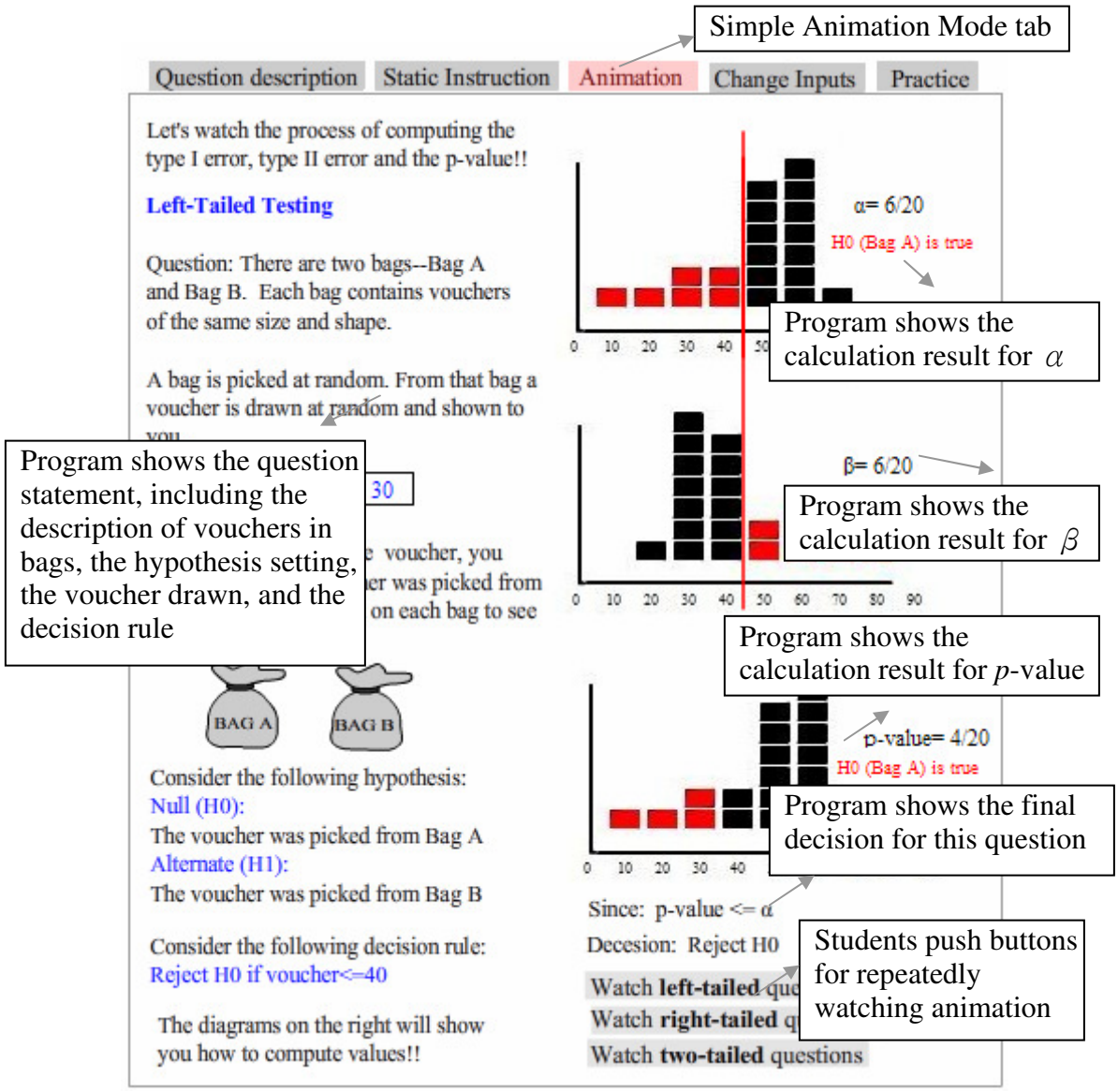

Figure 3.4 Program interface for the Simple Animation Mode

Change Inputs

In addition to allowing students to watch the animated procedure of calculating α, β, and *p*-value, the Change Inputs Mode also allowed learners to actively change the inputs/parameters under observation. Learners were allowed to select a H0 distribution, draw a random voucher from the bag, and set up a decision rule on the left side of the

window. After the student clicked on the "viewing calculation" button, the processes of calculating α, β, and the p-value were shown on the right side of the window with graphs and boxes. Students were allowed to stop, start and replay any animation for re-inspection. Figure 3.5 shows the Change Inputs Mode interface.

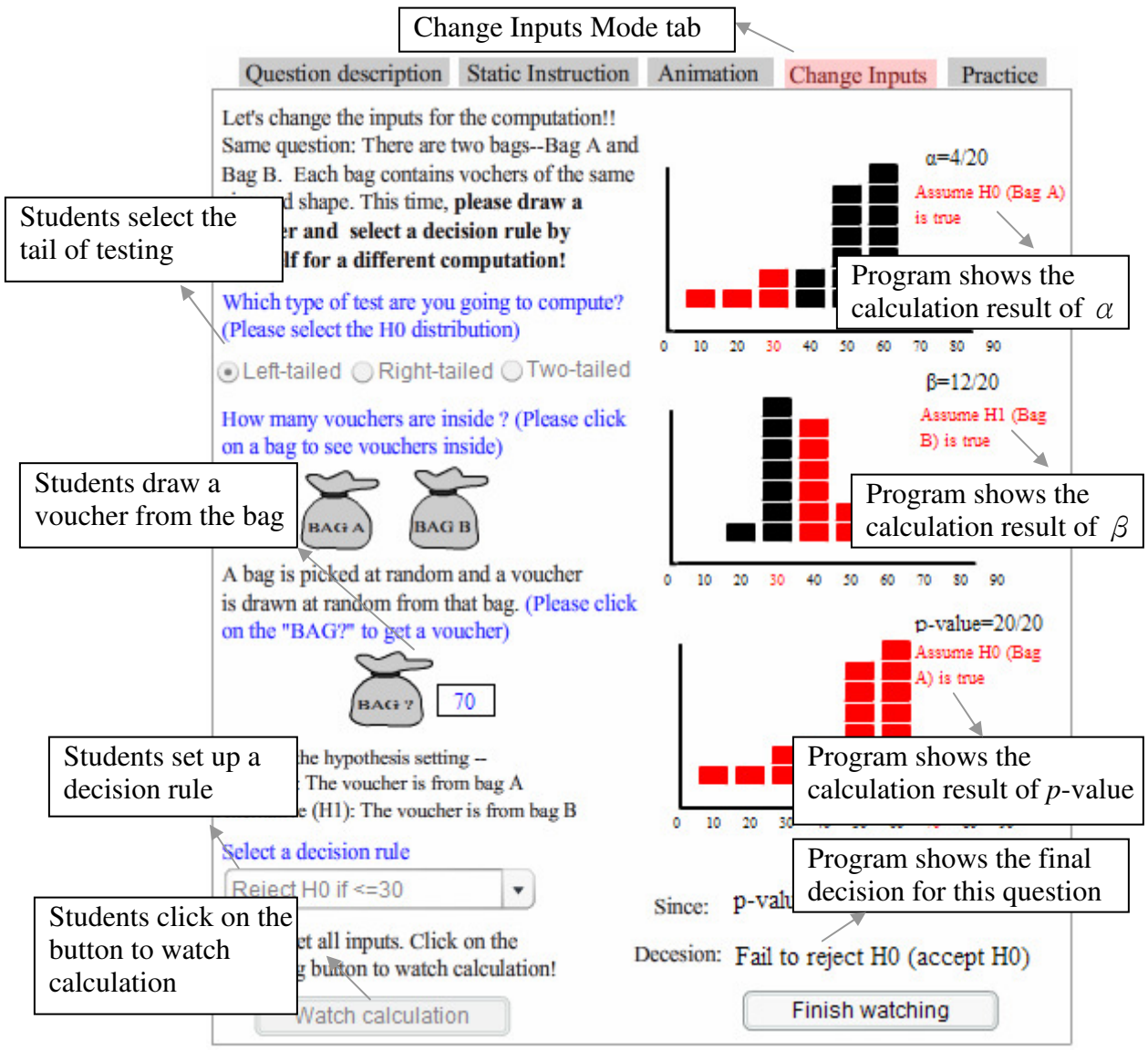

Figure 3.5 Program interface for the Change Inputs Mode

Figure 3.5 shows the calculation procedure of a one-sided test. The left side of the window allows students to alter the input. In this case, the H0 distribution is selected as a "left-tailed" distribution, the voucher drawn is set up as "70," and the decision rule is set up as "Reject H0 if value <= 30." The right side of the window shows the animation of calculation for α, β, and *p*-value. Three graphs are used to illustrate the calculation: The top one shows the calculation procedure for α, the black boxes represent the voucher values in the H0 bag, and the red boxes represent those vouchers which fall into the rejection region. In this case, while calculating α, there are 20 boxes (vouchers) in the bag, and only 4 fall in the rejection region. As a result, α equals four out of twenty (4/20). Similar calculation processes are animated in the middle and bottom graphs for β and *p*-value. After the calculations are finished, the testing conclusions are shown as texts in the bottom of the window. In this case, the program tells the students "since *p*-value $> \alpha$, we fail to reject H0." In this mode, learners can change any input for observation; however, they cannot obtain any feedback from the program.

Practice

The Practice Mode allowed students to calculate the α, β, and *p*-value themselves and receive feedback. There were 7 total questions for students to solve in each round, and students needed to use the graphs provided to do the calculation. In this mode, the program no longer showed students any calculation animation. Students were asked to select the correct distribution by using the drop-down menu for each question and clicking on the black boxes on the graphs to do the calculation. After calculating, students had to decide if the null hypothesis was rejected or accepted by selecting one item on the drop-down menu. When all the questions were answered, students could click on the "check my answer" button. The program showed students if they missed any

questions. The students could continue trying until they got all answers right. Figure 3.6 shows the Practice Mode interface.

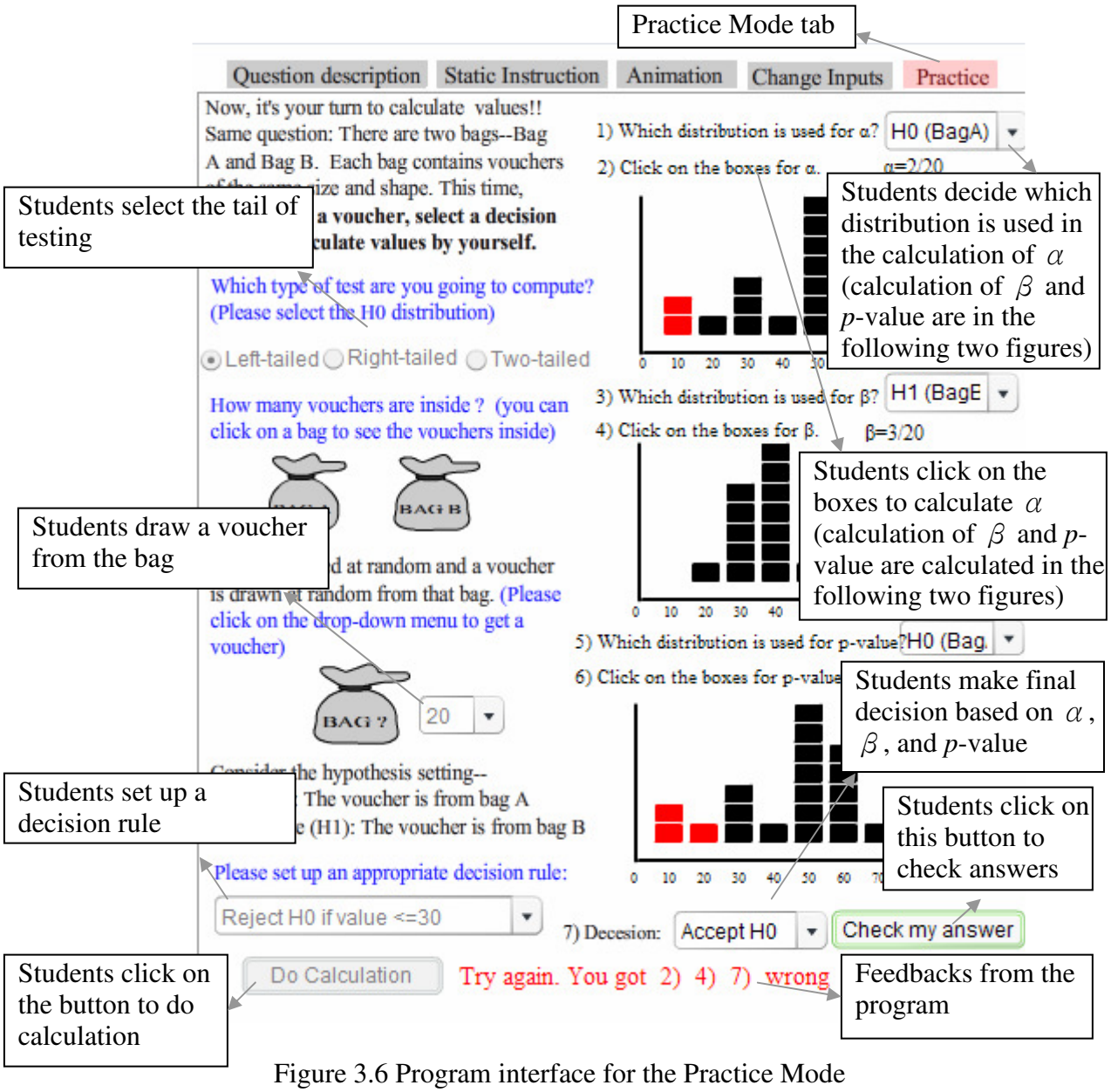

Figure 3.6 Program interface for the Practice Mode

Figure 3.6 shows an example of how a student practices questions in the Practice Mode. As with the Simple Animation Mode, the left side of the window allows the students to provide different inputs. In this case, the "left-tailed" distribution is selected

as the H0 distribution, the voucher value is set up as "20," and the decision rule is set up as "reject H0 if value <=30." After the student clicks on the "do calculation" button, 7 questions are shown on the right-side of the window. The first two questions ask students to select a distribution for α and click on the black boxes to mark those which fall in the rejection region. In this case, the student selects the "H0 distribution" for α, and clicks on the left two boxes to show the rejection region. As a result, the student's answer for α is two out of twenty (2/20). The calculation practice for β and the *p*-value are similar to that of α. For the value of β, in this case, "H1 distribution" is selected and the right three boxes are marked red. The student's answer for β is three out of twenty (3/20). For the *p*-value, in this case, "H0 distribution" is selected and the left three boxes are marked red. The student's answer for the *p*-value is three out of twenty (3/20). After comparing α with the *p*-value, the student chooses to accept H0. However, when the student checks his answers, the program tells him that he has three questions wrong: #2, #4, and #7. That is, the student has wrong answers for α, β, and the final decision.

In summary, this program provided animations which helped learners' development of mental models, and it was expected to improve student learning in introductory statistics. The following section discusses how the empirical study for this program was conducted.

RESEARCH DESIGN AND PROCEDURES

To understand how animation interactivity impacts student learning of hypothesis testing, an experiment was conducted with college-level students. The following sections describe the experiment.

Research questions and hypotheses

This research aimed to understand how animation interactivity influences learning outcomes. This study used Anderson and Krathwohl's (2001) cognitive taxonomy to divide knowledge delivered into three levels: remembering, understanding, and applying. Moreover, this study grouped the applying level into two sub-levels: Lower-level applying and higher-level applying, corresponding to procedure reasoning and integrated procedure reasoning in Garfield's reasoning model (2002). As a result, there were totally four levels of knowledge examined in this study: 1) Remembering, 2) understanding, 3) lower-level applying, and 4) higher-level applying.

The research questions and hypotheses tested in this study included:

1. Is there any difference in achievement improvement among students who use different interactive levels of an animation program?

H0-1: There is no difference in achievement improvement of *remembering* among groups who use different interactive levels of an animation program.

H0-2: There is no difference in achievement improvement of *understanding* among groups who use different interactive levels of an animation program.

H0-3: There is no difference in achievement improvement of *lower-level applying* among groups who use different interactive levels of an animation program.

H0-4: There is no difference in achievement improvement of *higher-level applying* among groups who use different interactive levels of an animation program.

2. Is there any difference in confidence improvement among students who use different interactive levels of an animation program?

H0-5: There is no difference in confidence improvement of *remembering* among groups who use different interactive levels of an animation program.

H0-6: There is no difference in confidence improvement of *understanding* among groups

who use different interactive levels of an animation program.

H0-7: There is no difference in confidence improvement of *lower-level applying* among groups who use different interactive levels of an animation program.

H0-8: There is no difference in confidence improvement of *higher-level applying* among groups who use different interactive levels of an animation program.

3. Is there any difference in program perception among students who use different interactive levels of an animation program?

H0-9: There is no difference in program perception among groups who use different interactive levels of an animation program.

Pilot study results and implications

Purpose and procedure of the pilot study

To test this particular animation program, a pilot study was conducted in a university in the third week of the 2009 spring semester (2-5-2009 to 2-14-2009). The purpose of the pilot study was to test the usability of the program and validity of the instruments used in the research. This section is only a summary of the pilot study results. Detailed analyses are attached in Appendix A.

The participants for the pilot study were 18 students enrolled in an undergraduate-level introductory statistics course. These students came from different colleges and departments with different majors. They were all novice statistics learners and this course was their first statistics related course in their college career. All participants joined the study voluntarily.

All students had an initial three-session lecture on the principles of hypothesis testing. After class, a website which included the surveys, tests, and the animation program was e-mailed to students. Students were asked to finish all the experimental

tasks at home using PC's or laptop computers. In this pilot study, students were not grouped by different treatments, and all students had the accessibility to all modes of the interactive animation. The experimental procedures are illustrated as follows:

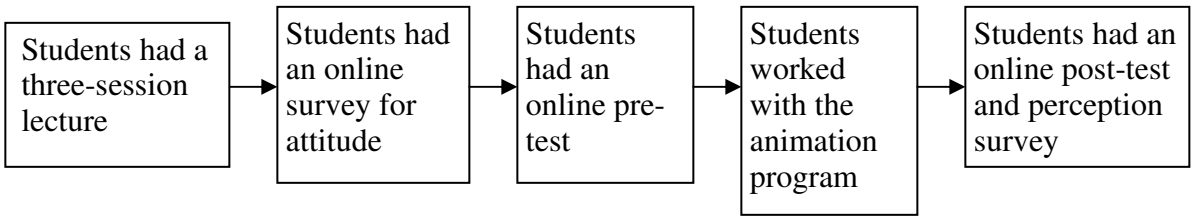

Figure 3.7 The experimental procedure of the pilot study

The pilot study used two surveys and two tests: The survey of attitudes toward statistics (SATS-36+) (Appendix B), the achievement and confidence pre-test (Appendix C), the achievement and confidence post-test (Appendix D), and the program perception survey(Appendix E). The following data were collected in this pilot study: learning attitude scores, achievement and confidence pre-test scores, achievement and confidence post-test scores, and program perception scores. Participants' Web log data was also collected. The details of these instruments are discussed in the following section.

Pilot study results

Before conducting pilot study data analysis, Cronbach's Alpha of internal consistency reliability test was run on all surveys and tests. This test measured the reliability of using these instruments (SATS-36+, achievement and confidence pre-test, achievement and confidence post-test, and program perception survey) for the participants in the study. The results indicated that all instruments were reliable. The Cronbach's Alpha coefficients for all instruments are listed as Table 3.1.

Table 3.1 Reliability coefficients for pilot study instruments

Instrument		Cronbach's Alpha
Learning attitude (SATS-36)	Affect	.77
	Cognitive Competence	.81
	Value	.87
	Perception of Difficulty	.73
	Interest	.89
	Effort	.72
Achievement pre-test (pilot)		.80
Achievement post-test (pilot)		.60
Program perception (pilot)		.91

Paired t-tests were conducted to examine if students improved after using the animation program. The results are recorded in Appendix A. In terms of achievement improvement, only students' scores on applying was significantly improved (t=-2.938, *p*=.009). Students' scores on remembering and understanding were not significantly enhanced. In terms of confidence improvement, only students' confidence levels on applying were significantly improved (t=-2.57, *p*=.020). Students' confidence levels on remembering and understanding were not significantly enhanced. Consequently, the pilot study indicated that significant improvement was only shown on students' applying ability.

In addition to gauging students' improvement using the instruments, another important purpose of the pilot study was to have students perform usability tests on the animation program. In the pilot study, students were told to contribute their thoughts and comments while using the animation program. All students' reflections were collected by the researcher for program revision and improvement.

Implication

The pilot study has several implications which suggest refinements for the methodology of the dissertation:

1. Increase of sample size

The testing results show that the use of the program only significantly improved students' learning achievement and confidence at the applying-level. Though students' knowledge at other levels (remembering and understanding) was not significantly improved, the *p*-values of these two levels were close to .05. The insignificance could have been due to the small sample size (only 18 participants only). It is very possible that these two levels of knowledge would be significantly improved by increasing the sample size to three or four times the current sample size.

2. Revision of pre-test and post-test questions

The reliability coefficients on the pilot pre-test and post-test were low. Also, the pilot study participants argued that the test procedure was redundant. As a result, the researcher removed repetitive items and re-organized remaining items to improve test quality.

3. Revision of program perception survey

Though the reliability coefficient on the program perception survey was high in the pilot study, there were validity and power issues since the survey lacked a rigorous validation process. Consequently, in the main study, the original survey will be replaced by Spaulding's (2007) multimedia perception survey which features a better validation process.

Subjects

The formal dissertation data collection was conducted in the 2009 fall semester. 127 students were recruited from the subject pool of a department at a large research

university in the southwest part of the U.S. Participants were randomly assigned to one of the four treatment groups. The details of the treatments are described as follows.

Treatment

In addition to a control group provided with only static materials, there were three groups with different levels of animation interactivity: 1) Animation with simple interactivity (Simple Animation Group), 2) animation with input manipulation (Change Inputs Group), and 3) animation with practice and feedback (Practice Group). The later group had all treatment provided in the previous group, plus it had a more advanced interactive function which was not possessed by the previous group.

Table 3.2 Treatment

Interactive levels	Program mode
Level 0: Static Group (control group) Traditional static Web-based instruction	Static Instruction Mode
Level1: Simple Animation Group Traditional static Web-based instruction + simple animation	Static Instruction Mode + Simple Animation Mode
Level2: Input Group Traditional static Web-based instruction + simple animation + animation with input manipulation	Static Instruction Mode + Simple Animation Mode + Change Inputs Mode
Level3: Practice Group Traditional static Web-based instruction + simple animation + animation with input manipulation + animation with practice and feedback	Static Instruction Mode + Simple Animation Mode + Change Inputs Mode + Practice Mode

Level0: Static group (control group)

In the Static group, students were allowed to use the Static Instruction Mode of the program. Students were provided with static Web pages on which information was similar to that of a traditional lecture. The program showed students the definitions and computations of type I error, type II error and *p*-values with a static pdf file. In this mode, students did not have the ability to exercise control, nor did they interact with any component of the program. This group comprised the control group in this experiment.

Level1: Simple Animation Group

In the Simple Animation Group, students could use the Static Instruction Mode as well as the Simple Animation Mode of the program. In addition to static Web pages, students were provided with the sequence-controlled animation. In this group, students had control over the sequences of these animations. For example, students could watch left-tailed, right-tailed or two-tailed animations in any order they prefer. The level of interaction afforded to this group was higher than that provided to the previous group.

Level2: Input Group

In the Input Group, students could change inputs to generate a different computing process for observation. In addition to the static Web pages and the simple animation, students were provided the opportunity to alter the inputs/variables to generate a visual representation. The level of interaction provided to this group was higher than that of the previous group.

Level3: Practice Group

In the Practice Group, students were allowed to answer questions and receive feedbacks from the program. In addition to the static Web pages, simple animation and input manipulation, students were provided with question practice opportunities. They

could repeatedly practice questions until they understood all the concepts. The level of interaction provided to this group was the highest among all of the groups.

Experimental procedure

All students gathered in a computer lab for approximately two hours to conduct the experiment. First, participants took an online survey of their learning attitudes. An online achievement and confidence pre-test followed the attitude survey. Participants were then randomly assigned to one of the four levels of interaction available in the animation program, and participated in sessions for using the animation program at their assigned levels of interaction. After they reviewed all the concepts in the unit, they could proceed to the online achievement and confidence post-test and perception survey. The following table shows the experimental procedures.

Table 3.3 The experimental procedure

Time	Tasks
3minutes	Subjects read and signed consent forms.
5-10 minutes	Subjects took the online learning attitude survey.
10-20 minutes	Subjects took the online achievement and confidence pre-test.
60-80 minutes	Subjects were randomly assigned to one of the treatment groups, and then used the animation program at their assigned interaction level.
10-20 minutes	Subjects took the achievement and confidence post-test.
1-3 minutes	Subjects took the perception survey.

Instruments and data collection

This study used two surveys and two tests: The survey of attitudes toward statistics (SATS-36+) (Appendix B), the achievement and confidence pre-test (Appendix F), the achievement and confidence post-test (Appendix G), and the perception survey for the program (Appendix H). In addition, the Web log data was recorded as an important source of data for analysis.

Attitude Survey: Survey of Attitudes toward Statistics (SATS-36 survey)

Candace Schau (1995) developed the Survey of Attitudes toward Statistics (SATS) to measure student attitudes while learning statistics (Appendix B). The SATS-36 version contained 36 items testing six attitude components in the inventory: Affect (six items; .80 to .89), cognitive competence (six items; .77 to .88), value (nine items; .74 to .90), and difficulty (seven items; .64 to .81). The reliability coefficients for interest and effort have not yet been released. However, this study obtained reliability coefficients for interest and efforts from the pilot study: Interest (four items; .89), efforts (four items; .72). SATS-36 survey is a seven-point scale survey, and the higher the student's scores, the more positive his attitude is. The following is a sample question on the survey:

1. I plan to complete all of my statistics assignments.

☐ 1 ☐ 2 ☐ 3 ☐ 4 ☐ 5 ☐ 6 ☐ 7

In the pilot study, the researcher used all six components as indicators for learning attitude. The results show that not all components work well to show student attitude. Consequently, in the formal data collection, only three of the six components were utilized: affect, interest, and perception of difficulty. These three sub-scales were the

most relevant to student learning motivation, and each student had three scores as their attitudes. The detail data items for SATS-36 are listed in the following table.

Table 3.4 SATS-36 scores

	Sub-scale	Number of questions
1	Affect (Scale) (Students' feelings concerning statistics)	6
2	Difficulty (Scale) (Students' attitudes on the difficulty of statistics as a subject)	7
3	Interest (Scale) (Students' level of individual interest in statistics)	4

Achievement and confidence pre-test and post-test

The achievement and confidence pre-test (Appendix F) and post-test (Appendix G) were both designed with 20 items, five items for a particular cognitive level on Bloom's revised model as mentioned in chapter 2 (remembering, understanding, lower-level applying, and higher-level applying). The tests composed of multiple choice questions. The questions tested students' knowledge of statistical hypothesis, including type I error (α), type II error (β) and the *p*-value. The two tests were composed of similar questions, comparable in difficulty and format. Students were also asked to provide their level of confidence for each question. The following shows some sample questions.

Question 1. (field 1-6)

There are two boxes: Box A and Box B.

Box A has 10 blue balls and 5 red balls.
Box B has 1 blue ball and 8 red balls.

A box is picked at random and a ball is drawn at random from that box and shown to you. Based on the color of the ball, you must decide if the ball was picked from Box A or Box B. Consider the following hypothesis:

Null H0: The ball was picked from Box A
Alternate H1: The ball was picked from Box B

Consider the following decision rule: Reject H0 if the ball is red.

Please calculate α, the type I error, and answer the following questions.

1.1.1 The above statement mentioned α (the type I error). Which one of the following is correct based on your understanding to α (the type I error)?

A) Give the statistical terminology that represents the chance of accepting H0 when in fact H1 is true.
B) Give the statistical terminology that represents the chance of rejecting H0 when in fact H0 is true.
C) None of above

1.1.2 How confident do you feel about your answer for 1.1.1?
A) Not confident at all
B) Not confident
C) Somewhat confident
D) Very confident

** This is a remembering-level question

The following table shows how the test questions are categorized based on cognitive ability level. Each student had eight scores on the achievement and confidence

tests (pre-test and post-test). The student's detailed data items for the tests are listed in the following table.

Table 3.5 The organization of achievement test questions

Revised Bloom's taxonomy	Garfield's reasoning model	Explanation	Question items	Total items
--	Idiosyncratic reasoning	--	--	--
Remembering	Verbal reasoning	Students are able to recall the definition or meaning of statistical concepts.	1.1.1 2.1.1 3.1.1 4.1.1 4.2.1	5
Understanding	Transitional reasoning	Students are able to explain, paraphrase or exemplify the statistical concepts.	1.2.1 2.2.1 3.2.1 4.3.1 4.4.1	5
Applying (Lower level)	Procedural reasoning	Students are able to solve application problems	1.3.1 2.3.1 3.3.1 4.5.1 4.6.1	5
Applying (Higher level)	Integrated process reasoning	Students are able to solve application problems with correct interpretation.	1.4.1 2.4.1 3.4.1 4.7.1 4.8.1	5

Table 3.6 Achievement and confidence test scores

	Sub-scale	Data item	Number of questions
1	Remembering	Achievement Score (Scale)	5
2		Confidence Score (Scale)	5
3	Understanding	Achievement Score (Scale)	5
4		Confidence Score (Scale)	5
5	Lower-level applying	Achievement Score (Scale)	5
6	(Procedural reasoning)	Confidence Score (Scale)	5
7	Higher-level-applying	Achievement Score (Scale)	5
8	(Integrated process reasoning)	Confidence Score (Scale)	5

Program perception survey

The program perception survey was revised from Spaulding's Survey for Student Learning and Assessment (2007). This scale was developed to understand student perception of the use of program (Appendix H). This survey contained four questions. It asked students on thoughts about the helpfulness of the program for statistics. The following is a sample question on the survey:

1. I think this multimedia program can help me learn.

☐ Not at all ☐ Less ☐ Somewhat ☐ More ☐ All the time

The student's detailed data items for the program perception survey are listed in the following table.

Table 3.7 Program perception scores

	Sub-scale	Number of questions
1	Perception score (Scale)	4

Web log data

After the student logged into the website, all his behavior was automatically recorded as Web log data. The Web log data was a supplementary data source in this study. It represented students' manipulation of the animation program. The Web log data included students' frequency of watching the animation, frequency of changing inputs, frequency of practice, and seconds spending in each mode. The following table shows these items in detail:

Table 3.8 Web log data

	Category	Data item
1	Frequency of watching the animation	The frequency of clicking on the animation button
2	Frequency of changing inputs	The frequency of clicking on the button for changing inputs
3	Frequency of doing practice	The frequency of clicking on the practice button
4	Seconds of playing the animation	The seconds spending in each mode

Data analysis

The hypotheses identified earlier were concerned with the significant differences in learning outcomes among the four groups. Several steps performed to test all of the hypotheses.

Tests for Cronbach's Alpha internal consistency reliability

Before testing any hypothesis, Cronbach's Alpha internal consistency reliability test was performed on all instruments. This was to calculate the internal reliabilities of all instruments used in the study.

Pre-test data analysis

The pre-test data analysis was conducted to show descriptive statistics on students' SATS-36 survey and achievement and confidence pre-test. This analysis helped to understand students' attitudes and prior knowledge.

Post-test data analysis

Multivariate procedure

A multivariate procedure was used to analyze the post-test data. The multivariate procedure aimed to answer research questions #1 and #2 (H0-1 to H0-8). The grouping variable was the animation interactivity, and the dependent variables were student achievement and confidence improvement. The student's achievement and confidence improvement were defined as the difference between one student's achievement and confidence pre-test scores and post-test scores. In other words, gain score analysis, instead of covariate analysis, was used in the study to examine students' improvement. Though either gain score analysis or covariate analysis can be used for pre-post data, this study picked gain score analysis as the main method based on Knapp and Schafer's argument in 2009:

> There is an important difference between the research question that is implied by the use of gain score analysis and the research question that underlies the use of ANCOVA. For the former, the question is: "What is the effect of the treatment on the change from pretest to posttest?" For the latter the question is: "What is the effect of the treatment on the posttest that is not predictable from the pretest?" (p.2)

The research question in this study was to understand students' learning improvement among different groups, closer to the former question in the above statement. Consequently, gain score analysis was the appropriate approach.

A multivariate procedure was used because it was designed to avoid the increase of Type I error which might result from the interactions among students' achievement scores and confidence levels. Consequently, a Multivariate Analysis of Variance (MANOVA) followed by Analyses of Variance (ANOVA) was conducted. The experiment was a one-way design with four groups (Static, Simple Animation, Input and Practice Group). After the MANOVA and ANOVA analysis, post hoc tests for group comparisons were conducted to determine where the significant differences came from.

Univariate Procedure

The univariate procedure aimed to answer research question #3 (H0-9). The grouping variable was the animation interactivity, and the dependant variable was student program perception. An Analysis of Variance (ANOVA) was used to examine students' program perception scores. Animation interactivity was the only independent variable in this study. In other words, the experiment was a one-way design with four groups (Static, Simple Animation, Input and Practice Group).

The following table shows a summary of research questions and methods for data analysis in the study.

Table 3.9 Summary of data analysis

Hypothesis	Method	Variables
H0-1: There is no difference in achievement improvement of *remembering* among groups who use different interactive levels of an animation program. H0-2: There is no difference in achievement improvement of *understanding* among groups who use different interactive levels of an animation program. H0-3: There is no difference in achievement improvement of *lower-level applying* among groups who use different interactive levels of an animation program. H0-4: There is no difference in achievement improvement of *higher-level applying* among groups who use different interactive levels of an animation program. H0-5: There is no difference in confidence improvement of *remembering* among groups who use different interactive levels of an animation program. H0-6: There is no difference in confidence improvement of *understanding* among groups who use different interactive levels of an animation program. H0-7: There is no difference in confidence improvement of *lower-level applying* among groups who use different	MANOVA	Grouping variable: Animation interactivity Dependent variables: Achievement improvement (4 sub-scores) and confidence improvement (4 sub-scores)

interactive levels of an animation program. H0-8: There is no difference in confidence improvement of student *higher-level applying* among groups who use different interactive levels of an animation program.		
H0-9: There is no difference in program perception among groups who use different interactive levels of an animation program.	ANOVA	Grouping variable: Animation interactivity Dependent variables: Program perception

Web log data analysis (Regression analysis, decision tree analysis, and case analysis)

The Web log data analysis included the following techniques: Regression analysis, regression tree analysis, and case analysis. Web log data was analyzed after all analyses were done, and it served as a supplementary source for explaining the testing results.

The regression analysis was performed to ascertain the relationship among animation manipulation, individual difference and learning. The predictors were student individual difference (including learning attitudes and prior knowledge) and their animation manipulation (including the frequency of watching animations, changing inputs and doing practice), and the predicted variable was their achievement post-test scores. The regression analysis would test if student animation manipulation or individual difference is a best predictor to their achievement outcomes.

In addition to the regression analysis, the classification and regression tree technique (CART) was used to determine interactions among variables. The regression

tree technique is a non-parametric technique that recursively partitions groups into smaller subgroups that maximally differ on a desired outcome (Witten & Frank, 2005). The regression tree is capable of accessing complex data patterns and higher order interactions, therefore helped to provide more valuable findings beyond the hypothesis testing results. The predictors and the predicted variables in the regression tree analysis were the same as those in the regression analysis.

Finally, the case analysis was performed to understand students' manipulation behaviors in detail. Seventeen students were randomly selected to participate in the case analysis (four students in the Static, Animation, and Input Group, and five students in the Practice Group). The researcher would report students' manipulation steps and then categorize students' behavior patterns. Student manipulation was expected to provide additional information to the hypothesis testing results.

SUMMARY

This chapter shows the interactive animation program and the research method used in the study. The animation program in this study aimed to help students learn the unit of hypothesis testing. Its design was based on Mayer's multimedia principles as well as Kristof and Satran's interactivity theories. This study used four instruments including both tests and surveys. The data items collected included student learning attitudes, achievement and confidence pre-test scores, achievement and confidence post-test scores, confidence levels, and program perception. Student manipulation of the animation program was recorded as Web log data. The data from the surveys and tests were analyzed by multivariate analysis (MANOVA) and univariate analysis (ANOVA). The Web log data was analyzed by regression analysis, regression tree analysis and the case analysis. The following section describes the results of the data analysis.

Chapter 4 Results

PARTICIPANT BACKGROUND

Of 127 participants, four participants' records were dropped from the data because these participants encountered some technical problems in the lab and did not finish the post-test. The final sample size was 123. Table 4.1 shows the distribution of subjects for each group.

Table 4.1 Distribution of subjects

Treatment group	Static	Simple Animation	Input	Practice	# of total participants
# of participants	30	30	31	32	123

All participants were undergraduate students recruited from the subject pool of a department at a large research university in the southwest part of the U.S. Among these participants, 95.28% had taken one statistics course and 4.72% had no statistics learning experiences.

RESULTS OF DATA ANALYSIS

The data analysis was divided into three steps. First, the reliability analysis was conducted. Then, the pre-test data was analyzed to show students' attitudes and prior knowledge. Finally, the post-test data and Web log data were analyzed with regard to the research questions.

Reliability of instrument

The reliability analysis was conducted on the attitude survey (SATS-36), program perception survey, and achievement and confidence tests (including pre-test and post-

test) to determine the internal consistency of students' responses to the test and survey items. The analysis for the SATS-36 indicated a Cronbach's alpha coefficient of .86 on the affect component, .72 on the difficulty perception component, and .90 on the interest component. For the achievement tests, coefficients of .68 and .74 were indicated on the pre-test and post-test respectively. For the confidence tests, coefficients of .94 and .97 were calculated on the pre-test and post-test respectively. The alpha coefficient for the program perception survey was .91. All of the results are summarized in Table 4.2. The coefficients were considered acceptable for these instruments.

Table 4.2 Reliability coefficients for main study instruments

Instrument		Cronbach's Alpha
SATS-36	Affect	.86
	Difficulty	.72
	Interest	.90
Achievement pre-test		.68
Pre-test confidence survey		.94
Achievement post-test		.74
Post-test confidence survey		.97
Program perception survey		.91

Pre-test data analysis

Descriptive statistics analyses were conducted on students' attitude scores (SATS-36), achievement pre-test scores, and confidence pre-test scores among the four groups. The purpose was to understand students' attitudes and prior knowledge.

SATS-36

Table 4.3 shows descriptive statistics for the SATS-36 scores for each group. Students' average attitude score on affect was 4.34, 3.72 on perception of difficulty, and 4.35 on interest.

Table 4.3 Descriptive statistics on SATS-36 scores for each group

Attitude component	Group	N	Mean	Std. Deviation	Min	Max
Affect	Static	30	4.3889	1.52397	1.17	7.00
	Simple Animation	30	4.3000	1.29203	1.67	6.83
	Input	31	4.4839	.95000	2.33	6.33
	Practice	32	4.1927	1.40475	1.00	6.83
	Total	123	4.3401	1.29775	1.00	7.00
Perception of difficulty	Static	30	3.8143	.99127	1.14	6.14
	Simple Animation	30	3.5429	.96117	1.86	5.14
	Input	31	3.8756	.66674	2.71	5.29
	Practice	32	3.6652	.77111	2.57	5.57
	Total	123	3.7247	.85464	1.14	6.14
Interest	Static	30	4.2083	1.58261	1.00	6.75
	Simple Animation	30	4.4667	1.56708	1.00	7.00
	Input	31	4.2742	1.39397	1.50	7.00
	Practice	32	4.4453	1.47798	1.00	7.00
	Total	123	4.3496	1.49135	1.00	7.00

* Attitude scores range from 1 to 7

Achievement and confidence pre-test

Table 4.4 summarizes the scores on students' achievement and confidence pre-test. The average achievement pre-test score on student remembering was 2.34, 2.48 on student understanding, 1.76 on student lower-level applying, and 2.11 on student higher-

level applying. The average confidence pre-test score on student remembering was 2.38, 2.07 on student understanding, 1.75 on student lower-level applying, and 1.95 on student higher-level applying.

Table 4.4 Descriptive statistics on achievement and confidence pre-test scores for each group

Pre-test Component	Group	N	Mean	Std. Deviation	Min	Max
Achievement scores on remembering	Static	30	2.33	1.269	0	4
	Simple Animation	30	2.57	1.357	0	5
	Input	31	2.52	1.208	0	5
	Practice	32	1.97	1.150	0	4
	Total	123	2.34	1.253	0	5
Achievement scores on understanding	Static	30	2.87	1.224	0	4
	Simple Animation	30	2.23	1.194	0	5
	Input	31	2.45	1.121	0	5
	Practice	32	2.38	1.008	0	5
	Total	123	2.48	1.148	0	5
Achievement scores on lower-level applying	Static	30	2.00	1.313	0	4
	Simple Animation	30	2.03	1.273	0	5
	Input	31	1.45	1.234	0	5
	Practice	32	1.59	1.292	0	4
	Total	123	1.76	1.287	0	5
Achievement scores on higher-level applying	Static	30	2.23	1.278	0	5
	Simple Animation	30	2.17	1.440	0	5
	Input	31	2.00	.894	1	4
	Practice	32	2.03	1.092	0	4
	Total	123	2.11	1.179	0	5
Confidence scores on	Static	30	2.5733	1.13348	1.00	4.60

remembering	Simple Animation	30	2.2267	1.03555	1.00	5.00
	Input	31	2.4452	.78606	1.00	4.00
	Practice	32	2.2937	.90302	1.00	4.40
	Total	123	2.3837	.96821	1.00	5.00
Confidence scores on understanding	Static	30	2.2067	.87649	1.00	4.00
	Simple Animation	30	1.9733	.88938	1.00	4.40
	Input	31	2.1290	.75815	1.00	4.00
	Practice	32	1.9875	.63640	1.00	3.60
	Total	123	2.0732	.79021	1.00	4.40
Confidence scores on lower-level applying	Static	30	1.8067	.70756	1.00	4.00
	Simple Animation	30	1.5333	.57135	1.00	3.20
	Input	31	1.7871	.70604	1.00	3.20
	Practice	32	1.8687	.70135	1.00	3.40
	Total	123	1.7512	.67864	1.00	4.00
Confidence scores on higher-level applying	Static	30	2.1000	.83004	1.00	4.00
	Simple Animation	30	1.7733	.73669	1.00	3.80
	Input	31	1.9871	.70793	1.00	4.00
	Practice	32	1.9375	.65537	1.00	3.40
	Total	123	1.9496	.73400	1.00	4.00

*Points for achievement scores range from 0 to 5 / Points for confidence scores range from 1 to 5

*Achievement improvement ranges from 0 to 5 / Confidence improvement ranges from 0 to 4

MANOVA and ANOVAs were conducted to see if there were significant differences on attitude, achievement and confidence among groups. The results showed there were no significant differences among groups in terms these three factors (p>.05). Consequently, this study ensured that no significant differences existed among groups. In other words, the study met the requirement of equal grouping and no covariate variables were needed to be set.

Post-test data analysis

The post-test data analysis included a MANOVA on student achievement and confidence improvement, an ANOVA on student program perception, and Web log data analysis. Before going to the procedure of MANOVA, students' engagement of the animation program are shown as the following table.

Table 4.5 Total seconds spent for each group

Group	Mean	N	Std. Deviation	Range
Static	349.3333	30	369.39618	1119.00
Simple Animation	525.9667	30	472.34868	1701.00
Input	669.1290	31	441.01948	1837.00
Practice	797.2500	32	628.11284	2448.00
Total	589.5447	123	511.52186	2458.00

One-Way Multivariate Analysis of Variance: Analysis of Achievement and Confidence Scores

Table 4.6 shows descriptive statistics on the achievement and confidence post-test scores for each group, and Table 4.7 shows descriptive statistics on student achievement and confidence improvement for each group. Student achievement improvement was defined as the difference between one student's achievement pre-test and post-test scores. Similarly, student confidence improvement was defined as the difference between one student's confidence pre-test and post-test scores.

Table 4.6 Descriptive statistics on achievement and confidence post-test scores for each group

Component	Group	N	Mean	Std. Deviation	Min	Max
Achievement scores on remembering	Static	30	2.97	1.326	0	5
	Simple Animation	30	3.30	1.236	0	5
	Input	31	3.55	1.121	1	5

	Practice	32	2.97	1.332	0	4
	Total	123	3.20	1.265	0	5
Achievement scores on understanding	Static	30	2.93	1.484	0	5
	Simple Animation	30	3.57	1.073	2	5
	Input	31	3.61	1.334	1	5
	Practice	32	3.72	1.397	0	5
	Total	123	3.46	1.351	0	5
Achievement scores on lower-level applying	Static	30	2.63	1.542	0	5
	Simple Animation	30	2.60	1.354	0	5
	Input	31	2.94	1.365	0	5
	Practice	32	2.94	1.294	0	5
	Total	123	2.78	1.382	0	5
Achievement scores on higher-level applying	Static	30	2.53	1.306	0	5
	Simple Animation	30	2.90	1.094	1	5
	Input	31	2.65	1.450	0	5
	Practice	32	3.06	1.134	0	4
	Total	123	2.79	1.256	0	5
Confidence scores on remembering	Static	30	3.4133	1.32866	1.00	5.00
	Simple Animation	30	3.6867	1.17289	1.00	5.00
	Input	31	3.7355	.94005	1.00	5.00
	Practice	32	3.6000	1.03363	1.00	5.00
	Total	123	3.6098	1.11821	1.00	5.00
Confidence scores on understanding	Static	30	3.0267	1.22698	1.00	5.00
	Simple Animation	30	3.1200	1.02197	1.00	5.00
	Input	31	3.2258	.85439	1.00	4.40
	Practice	32	3.2063	1.01025	1.00	5.00
	Total	123	3.1463	1.02528	1.00	5.00
Confidence scores on lower-level applying	Static	30	2.7400	1.19787	1.00	5.00
	Simple Animation	30	2.8800	1.11985	1.00	5.00
	Input	31	3.1161	.87143	1.00	4.60
	Practice	32	2.9375	1.01529	1.00	5.00
	Total	123	2.9203	1.05169	1.00	5.00
Confidence scores on	Static	30	2.5733	1.11136	1.00	4.60

higher-level applying	Simple Animation	30	2.6267	1.15875	1.00	4.60
	Input	31	2.7806	.75252	1.00	3.80
	Practice	32	2.7937	.96718	1.00	4.60
	Total	123	2.6959	.99946	1.00	4.60

*Points for achievement scores range from 0 to 5 / Points for confidence scores range from 1 to 5

*Achievement improvement ranges from 0 to 5 / Confidence improvement ranges from 0 to 4

Table 4.7 Descriptive statistics on achievement and confidence improvement for each group

Component	Group	N	Mean	Std. Deviation	Min	Max
Achievement improvement on remembering	Static	30	.63	1.564	-3	4
	Simple Animation	30	.73	1.507	-2	4
	Input	31	1.03	1.426	-3	4
	Practice	32	1.00	1.566	-4	3
	Total	123	.85	1.508	-4	4
Achievement improvement on understanding	Static	30	.07	1.507	-2	3
	Simple Animation	30	1.33	1.605	-2	4
	Input	31	1.16	1.655	-2	4
	Practice	32	1.34	1.677	-2	5
	Total	123	.98	1.679	-2	5
Achievement improvement on lower-level applying	Static	30	.63	1.497	-3	3
	Simple Animation	30	.57	1.633	-3	4
	Input	31	1.48	1.568	-4	5
	Practice	32	1.34	1.558	-2	5
	Total	123	1.02	1.599	-4	5
Achievement improvement on higher-level applying	Static	30	.30	1.418	-2	3
	Simple Animation	30	.73	1.337	-2	4
	Input	31	.65	1.305	-2	3
	Practice	32	1.03	1.694	-2	4
	Total	123	.68	1.456	-2	4
Confidence	Static	30	.8400	1.01084	-1.00	2.80

improvement on remembering	Simple Animation	30	1.4600	1.10066	-.60	3.40
	Input	31	1.2903	1.01828	-.60	3.20
	Practice	32	1.3063	.93116	-.20	3.20
	Total	123	1.2260	1.02907	-1.00	3.40
Confidence improvement on understanding	Static	30	.8200	.91779	-.40	2.80
	Simple Animation	30	1.1467	1.01429	-1.20	3.20
	Input	31	1.0968	.90018	-.60	2.80
	Practice	32	1.2188	.81178	.00	2.80
	Total	123	1.0732	.91339	-1.20	3.20
Confidence improvement on lower-level applying	Static	30	.9333	1.09838	-.60	3.40
	Simple Animation	30	1.3467	1.10414	-.40	3.40
	Input	31	1.3290	1.11779	-.60	3.40
	Practice	32	1.0688	1.01931	-.40	3.20
	Total	123	1.1691	1.08553	-.60	3.40
Confidence improvement on higher-level applying	Static	30	.4733	.77812	-1.20	2.60
	Simple Animation	30	.8533	.92912	-.20	3.40
	Input	31	.7935	.89664	-.80	2.60
	Practice	32	.8563	.67439	-.20	2.80
	Total	123	.7463	.82901	-1.20	3.40

*Points for achievement scores range from 0 to 5 / Points for confidence scores range from 1 to 5

*Achievement improvement ranges from 0 to 5 / Confidence improvement ranges from 0 to 4

Student achievement and confidence improvement were analyzed by the one-way multivariate analysis of variance (MANOVA). The independent variable was animation interactivity and the dependent variables included four achievement scores and four confidence scores. It was important to use this procedure because of the possibility of interactions among student achievement and confidence. This interaction might have resulted in an increased Type I error rate. As a result, a MANOVA providing a better control for Type I error was selected to analyze the post-test data. In addition, an alpha-level of .05 was used for all analyses in the study. Therefore the assumption of

homogeneity of variance was satisfied in this analysis. The result of the MANOVA is summarized in Table 4.8.

Table 4.8 MANOVA on learning improvement

Effect		Value	F	Hypothesis df	Error df	Sig.
Intercept	Pillai's Trace	.669	28.235(a)	8.000	112.000	.000
	Wilks' Lambda	.331	28.235(a)	8.000	112.000	.000
	Hotelling's Trace	2.017	28.235(a)	8.000	112.000	.000
	Roy's Largest Root	2.017	28.235(a)	8.000	112.000	.000
Group	Pillai's Trace	.275	1.436	24.000	342.000	.087
	Wilks' Lambda	.747	1.433	24.000	325.435	.089
	Hotelling's Trace	.309	1.427	24.000	332.000	.091
	Roy's Largest Root	.159	2.263(b)	8.000	114.000	.028

a Exact statistic
b The statistic is an upper bound on F that yields a lower bound on the significance level.
c Design: Intercept+group

The *p*-values of Levene's Test for Equality of Variances for the tests were .685, .644, .686, .571, .766, .802, .862 and .357 (to the eight variables respectively). Therefore the assumption of homogeneity of variance was satisfied in this analysis.

Among all multivariate criteria, there was an overall significant difference in student achievement and confidence improvement among groups when Roy's Largest Root is used (p=.028).

Then ANOVAs were performed following the MANOVA procedure. In terms of student achievement improvement, there were significant differences among groups in the understanding ($F(3,119)=4.354$, p=.006) and the lower-level applying ($F(3,119)=2.815$, p=.042) (Table 4.8). In terms of student confidence improvement, there was no significant difference among groups (Table 4.9).

Table 4.9 Following-up ANOVAs on learning improvement

Source	Dependent Variable	Type III Sum of Squares	Df	Mean Square	F	Sig.
Corrected Model	Achievement improvement on remembering	3.565(a)	3	1.188	.516	.672
	Achievement improvement on understanding	34.022(b)	3	11.341	4.354	.006
	Achievement improvement on lower-level applying	20.673(c)	3	6.891	2.815	.042
	Achievement improvement on higher-level applying	8.402(d)	3	2.801	1.332	.267
	Confidence improvement on remembering	6.447(e)	3	2.149	2.083	.106
	Confidence improvement on understanding	2.780(f)	3	.927	1.114	.346
	Confidence improvement on lower-level applying	3.729(g)	3	1.243	1.056	.371
	Confidence improvement on higher-level applying	3.035(h)	3	1.012	1.490	.221
Intercept	Achievement improvement on remembering	88.748	1	88.748	38.572	.000
	Achievement improvement on understanding	117.145	1	117.145	44.977	.000
	Achievement improvement on lower-level applying	124.615	1	124.615	50.908	.000
	Achievement improvement on higher-level applying	56.407	1	56.407	26.825	.000
	Confidence improvement on remembering	184.186	1	184.186	178.560	.000
	Confidence improvement on understanding	140.866	1	140.866	169.322	.000
	Confidence improvement on lower-level applying	168.094	1	168.094	142.846	.000
	Confidence improvement on higher-level applying	68.057	1	68.057	100.220	.000
Group	Achievement improvement on remembering	3.565	3	1.188	.516	.672
	Achievement improvement on understanding	**34.022**	**3**	**11.341**	**4.354**	**.006 (*)**
	Achievement improvement on lower-level applying	**20.673**	**3**	**6.891**	**2.815**	**.042 (*)**
	Achievement improvement on higher-level applying	8.402	3	2.801	1.332	.267
	Confidence improvement on remembering	6.447	3	2.149	2.083	.106
	Confidence improvement on understanding	2.780	3	.927	1.114	.346
	Confidence improvement on lower-level applying	3.729	3	1.243	1.056	.371
	Confidence improvement on higher-level applying	3.035	3	1.012	1.490	.221
Error	Achievement improvement on remembering	273.801	119	2.301		
	Achievement improvement on understanding	309.946	119	2.605		
	Achievement improvement on lower-level applying	291.294	119	2.448		
	Achievement improvement on higher-level applying	250.232	119	2.103		
	Confidence improvement on remembering	122.750	119	1.032		
	Confidence improvement on understanding	99.001	119	.832		
	Confidence improvement on lower-level applying	140.034	119	1.177		

	Confidence improvement on higher-level applying	80.811	119	.679		
Total	Achievement improvement on remembering	367.000	123			
	Achievement improvement on understanding	463.000	123			
	Achievement improvement on lower-level applying	439.000	123			
	Achievement improvement on higher-level applying	316.000	123			
	Confidence improvement on remembering	314.080	123			
	Confidence improvement on understanding	243.440	123			
	Confidence improvement on lower-level applying	311.880	123			
	Confidence improvement on higher-level applying	152.360	123			
Corrected Total	Achievement improvement on remembering	277.366	122			
	Achievement improvement on understanding	343.967	122			
	Achievement improvement on lower-level applying	311.967	122			
	Achievement improvement on higher-level applying	258.634	122			
	Confidence improvement on remembering	129.197	122			
	Confidence improvement on understanding	101.781	122			
	Confidence improvement on lower-level applying	143.763	122			
	Confidence improvement on higher-level applying	83.846	122			

* $p < .05$

Tukey HSD post hoc comparisons were then conducted to find the source of the difference. Table 4.10 shows the results of post hoc comparison. The results indicated that:

1) Student understanding improvement in the Simple Animation, Input and Practice Group was significantly better than the Static Group.
2) Student understanding improvement in the Practice Group was not significantly best.

Table 4.10 Tukey HSD post-hoc comparisons on learning improvement

Dependent Variable	(I) Group	(J) Group	Mean Difference (I-J)	Sig.
Achievement improvement on understanding	Static	Simple Animation	-1.27	.015(*)
		Input	-1.09	.045(*)
		Practice	-1.28	.012(*)
	Simple Animation	Static	1.27	.015(*)
		Input	.17	.976
		Practice	-.01	1.000
	Input	Static	1.09	.045(*)
		Simple Animation	-.17	.976
		Practice	-.18	.970
	Practice	Static	1.28	.012(*)
		Simple Animation	.01	1.000
		Input	.18	.970

Based on observed means.
* The mean difference is significant at the .05 level.

For more conservative post hoc comparisons (Bonferroni correction for post hoc comparisons), please see Appendix J. For details of the effect size of the animation interactivity on achievement and confidence improvement, please see Appendix K.

Analysis of program perception

The program perception analysis by ANOVA was conducted to understand if there was any significant difference in program perception among the four groups. Table 4.11 shows the descriptive statistics on student program perception for each group, and Table 4.12 shows the testing results. The results indicated that there was no significant difference in program perception among the four groups.

The *p*-value of Levene's Test for Equality of Variances for the test was .096. Therefore the assumption of homogeneity of variance was satisfied in this analysis.

Table 4.11 Descriptive statistics on program perception for each group

Group	N	Mean	Std. Deviation	Min	Max
Static	30	3.2917	1.12970	1.00	5.00
Simple Animation	30	3.4250	1.25799	1.00	5.00
Input	31	3.4274	1.04322	1.00	5.00
Practice	32	3.7656	.86821	1.00	5.00
Total	123	3.4817	1.08178	1.00	5.00

*Program perception scores range from 1 to 5

Table 4.12 ANOVA on program perception

	Sum of Squares	Df	Mean Square	F	Sig.
Between Groups	61.613	3	20.538	1.100	.352
Within Groups	2222.729	119	18.678		
Total	2284.341	122			

For details of the effect size of the animation interactivity on student program perception, please see Appendix K.

Summary of hypotheses testing results

The results indicated that the increase of animation interactivity could enhance student achievement improvement on understanding and lower-level applying, but not on student remembering and higher-level applying. In addition, student confidence

improvement and program perception were not enhanced by the increase of animation interactivity. Table 4.13 shows the summary of the above hypothesis testing results.

Table 4.13 Summary of hypotheses testing results

Hypothesis	Results
H0-1: There is no difference in achievement improvement of *remembering* among groups who use different interactive levels of an animation program.	Accept
H0-2: There is no difference in achievement improvement of *understanding* among groups who use different interactive levels of an animation program.	Reject
H0-3: There is no difference in achievement improvement of *lower-level applying* among groups who use different interactive levels of an animation program.	Reject
H0-4: There is no difference in achievement improvement of *higher-level applying* among groups who use different interactive levels of an animation program.	Accept
H0-5: There is no difference in confidence improvement of *remembering* among groups who use different interactive levels of an animation program.	Accept
H0-6: There is no difference in confidence improvement of	Accept

understanding among groups who use different interactive levels of an animation program.	
H0-7: There is no difference in confidence improvement of *lower-level applying* among groups who use different interactive levels of an animation program.	Accept
H0-8: There is no difference in confidence improvement of *higher-level applying* among students who use different interactive levels of an animation program.	Accept
H0-9: There is no difference in program perception among students who use different interactive levels of an animation program.	Accept

The results showed that the increase of animation interactivity did not enhance student learning improvement at all levels of knowledge. To better understand the above testing results, the Web log data analysis was then conducted.

Web log data analysis

Multiple regression and regression tree analysis

The purpose of the Web log data analysis was to understand the relationship among student animation manipulation, individual differences and achievement post-test scores. The analysis allowed the researcher to explore why the increase of animation interactivity was not always helpful. The data were first analyzed with multiple linear regression, using student animation manipulation (including the frequency of watching an animation, the frequency of changing inputs, and the frequency of doing practice) and individual differences (including student prior knowledge, affect, perception of difficulty and interest) as predictors, and student achievement post-test score as predicted variables.

Table 4.14 shows the R square of the regression, Table 4.15 shows the significance level of this regression model, and Table 4.16 shows the regression coefficients.

Table 4.14 R square of the regression analysis

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	.521(a)	.272	.228	3.385

a Predictors: (Constant) , prior knowledge(pre-test), affect, interest, perception of difficulty, watch_frequency, change_frequency, practice_frequency
b Dependent Variable: posttest

Table 4.15 Significance level of the regression analysis

Model		Sum of Squares	Df	Mean Square	F	Sig.
1	Regression	492.063	7	70.295	6.135	.000(a)(*)
	Residual	1317.563	115	11.457		
	Total	1809.626	122			

a Predictors: (Constant) , prior knowledge(pre-test), affect, interest, perception of difficulty, watch_frequency, change_frequency, practice_frequency
b Dependent Variable: posttest
* $p < .05$

Table 4.16 Regression coefficients

	Model	Unstandardized Coefficients	Standardized Coefficients (Beta)	t	Sig.
1	(Constant)	6.890		3.922	.000
	Prior knowledge (pre-test scores)	.387	.297	3.529	.001(*)
	Affect	.267	.090	.742	.460
	Interest	.638	.247	2.398	.018(*)
	Perception of difficulty	-.732	-.162	-1.688	.094
	Frequency of watching animaiton	.282	.144	1.714	.089
	Frequency of changing inputs	.142	.099	.891	.375
	Frequency of practicing	.010	.016	.149	.882

a Dependent Variable: post-test
* $p < .05$

The tables show that the regression coefficients of prior knowledge and interest were significant at the .05 level. This result could be interpreted as follows:

1. Increasing the student's prior knowledge by 1 standard deviation would make his achievement post-test scores increase by .297 standard deviation while holding other variables in control. Moreover, the coefficient of students' prior knowledge was significant at the .05 level in the regression analysis (p=.001).
2. Increasing the student's interest towards statistics by 1 standard deviation would make his achievement post-test scores increase by .247 standard deviation while holding other variables in control. In addition, the coefficient of student interest was significant at the .05 level in the regression analysis (p=.018).

The model R-square was .228 ($F(7,115)=6.135$, $p<.000$), reflecting the overall strength of the relationship among predictors and predicted variables. In other words, 27% of variance was explained by those seven variables.

Interestingly, the study found that the two most important predictors of students' achievement post-test scores were student prior knowledge and interest, instead of animation manipulation. Then the classification and regression tree (CART) technique was used to understand the interactions among the predictors and the predicted variable. In this tree analysis, the independent variables were student animation manipulation (including frequency of watching animations, changing inputs and doing practices) and individual differences (including affect, interest, difficulty, prior knowledge). The dependent variable was student achievement post-test performance, which was defined as a categorical variable with two values: below average and above average. In other words, the regression tree analysis aimed to predict how a student would perform on the post-test

using his animation manipulation and individual differences. The maximum tree depth was set as five (default). Figure 4.1 shows the regression tree. Group 1 was the group below average, and group 2 was the group above average.

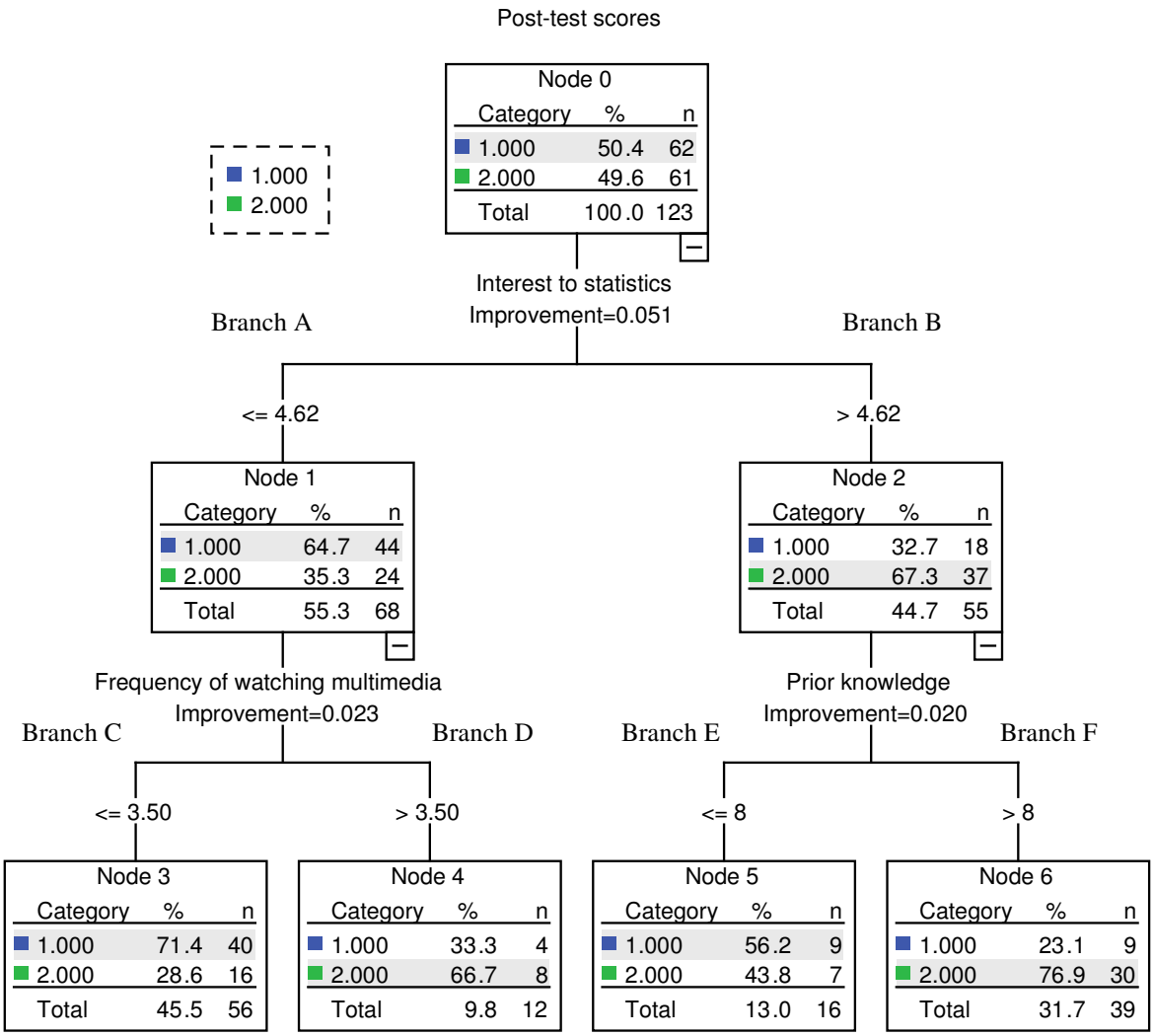

Figure 4.1 Regression tree: Individual differences and animation manipulation

The sample size was 123. The percentage for group 1 (below average) was 50.4%, and for group 2 (above average) was 49.6%. The tree was sorted such that the most

significant splitter, Interest, was at the top. The students with low interest (<=4.62) went to the left (Branch A) and those with high interest (>4.62) went to the right (Branch B). When student interest was 1ow (Branch A), over 64% of the students had low post-test scores (below average); when student interest was high (Branch B), then over 67% of the students had high post-test scores (above average).

Under Branch A, Frequency of Watching Animation was the next splitter. The students with low frequency of watching animation (<=3.50) went to the left (Branch C) and those with high frequency of watching animation (>3.50) went to the right (Branch D). When student frequency of watching animation was 1ow (Branch C), over 71% of the students had low post-test scores (below average); when students' frequency of watching animation was high (Branch D), then over 66% of the students had high post-test scores (above average).

Under Branch B, Prior Knowledge was the next splitter. The students with low prior knowledge (<=8) went to the left (Branch E) and those with high prior knowledge (>8) went to the right (Branch F). When student prior knowledge was 1ow (Branch E), over 56% of the students had low post-test scores (below average); when student prior knowledge was high (Branch F), then over 76% of the students had high post-test scores (above average).

The group which had the best outcome was Node 6. Node 6 was composed of students with high interest and high prior knowledge; 76.9% of Node 6 students achieved high post-test performance (above average). The group which had the second highest outcome was Node 4. Node 4 was composed of students with low interest but high frequency of watching animation, and 66.7% of Node 4 students achieved high post-test performance. The group which had the third highest outcome was Node 5. Node 5 was composed of students with high interest but low prior knowledge, and 43.8% of Node 5

students achieved high post-test performance. The group which had the worst outcome was Node 3. Node 3 was composed of students with low interest and low frequency of watching animation, and only 28.6% of Node 3 students achieved high post-test performance.

The tree analysis supported the findings in the regression analysis. It showed that student prior knowledge and interest were important predictors of student learning outcome. Frequency of watching an animation helps for some of the students (e.g., low-interest students) but not to all. To better understand the impact of animation manipulation on learning, this study then picked 17 students for case analysis, thoroughly examining students' use of the program.

Case analysis

The hypothesis testing result showed that the increase of animation interactivity benefits learning at certain levels (understanding and lower-level applying). This might be because student learning attitudes, prior knowledge, and program manipulation mediated the animation impact. The regression tree analysis showed that there were interactions on student achievement post-test scores among these three factors. To better understand this issue, this study picked 4-5 students in each group for case analysis (Table 4.17). The case analysis used the Web log records to describe how students went through this animation program and then generated behavior patterns for comparison. To follow the privacy and confidentiality policy requested by the university, participants in the study were recorded and reported by using pseudo-names. The analysis was expected to provide in-depth information on the relationship among individual student differences, animation manipulation and achievement post-test scores.

Table 4.17 Case selection

Group	Case Selection		
	Low prior knowledge (pre-test achievement scores<10)	High prior knowledge (pre-test achievement scores>=10)	Total case
Static Group	2	2	4
Simple Animation Group	2	2	4
Input Group	2	2	4
Practice Group	2	3	5
Total case	8	9	17

1. Static Group

Students assigned to the Static Group received only the static instruction. Students could switch between the Question Statement Page and the Static Instruction Mode. This study picked four cases for analysis: Allen, Jenny, Wendy, and Jeff. Their manipulation pattern in the Static Group was very simple. That is, they always switched between the Question Statement Page and the Static Instruction Mode:

Pattern 1: Question Statement Page-> Static Instruction Mode (repeated)

Case1: Allen (#179)- low prior knowledge student

Allen was a student with low prior knowledge and a very negative attitude towards statistics (affect score =1.17, interest score =1.00, perception of difficulty score =3.14, achievement pre-test score= 5). Due to his negative attitude, he was impatient while playing the program. Allen spent a total of 127 seconds on the program. The following is how he used the program:

1) He first spent 42 seconds reading the Question Statement Page.
2) He then spent 61 seconds on the Static Instruction Mode.
3) He finally spent 24 seconds reading the Question Statement Page again.

Allen switched 1.5 rounds between the Question Statement Page and the Static Instruction Mode (step 1-3). He was engaged in the program only for a very short period; consequently his achievement post-test showed little improvement (post-test score=6). His perception towards the program was also very low (program perception score=1). In other words, he did not consider this program helpful.

Case2: Jenny (#136)- low prior knowledge student

Jenny was another student with low prior knowledge and a negative attitude towards statistics (affect score =2.83, interest score =1.75, perception of difficulty score =2.57, achievement pre-test score= 8). Though Jenny had an attitude and prior knowledge similar to Allen's, she spent more time exploring the program. She spent a total of 1032 seconds on the program. The following is how she used the program:

1) She first spent 68 seconds reading the Question Statement Page.
2) She then spent 480 seconds on the Static Instruction Mode.
3) She went back to the Question Statement Page for 9 seconds.
4) She spent 12 seconds on the Static Instruction Mode.
5) She spent 69 seconds on the Question Statement Page.
6) She spent 101 seconds on the Static Instruction Mode.

7) She spent 9 seconds on the Question Statement Page.

8) She spent 37 seconds on the Static Instruction Mode.

9) She spent 4 seconds on the Question Statement Page.

10) She spent 149 seconds on the Static Instruction Mode.

11) She spent 33 seconds on the Question Statement Page.

12) She spent 20 seconds on the Static Instruction Mode.

13) She spent 4 seconds on the Question Statement Page.

14) She spent 37 seconds on the Static Instruction Mode.

Jenny switched 7 rounds between the Question Statement Page and the Static Instruction Mode (step 1-14). Her perception of the program was very positive (program perception score=5.0), which means she thought this program was helpful for her learning. Her achievement post-test showed substantial improvement (post-test score =15). Jenny's post-test score was one of the highest among all Static Group students.

Case 3: Wendy (#64)- high prior knowledge student

In contrast to Allen and Jenny, Wendy was a student with high prior knowledge and a positive attitude towards statistics (affect score =7.00, interest score =6.00, perception of difficulty score =3.14, achievement pre-test score= 12). Wendy spent a total of 854 seconds on the program. The following is how she used the program:

1) She first spent 123 seconds reading the Question Statement Page.

2) She then spent 731 seconds on the Static Instruction Mode.

Though Wendy switched only 1 round between the Question Statement Page and the Static Instruction Mode for learning (step 1-2), the time she stayed on each mode was quite long. Wendy's achievement post-test showed substantial improvement (post-test score=18). However, her perception to the program was negative (program perception score=2.25), which means she thought this program not very helpful for her learning.

Case 4: Jeff (#156)- high prior knowledge student

Jeff was another student with high prior knowledge and a positive attitude towards statistics (affect score =4.83, interest score =6.00, perception of difficulty score =5.86, achievement pre-test score= 10). Jeff spent a total of 1,014 seconds on the program. The following is how he used the program:

1) He first spent 136 seconds reading the Question Statement Page.
2) He then spent 143 seconds on the Static Instruction Mode.
3) He went back to the Question Statement Page for 32 seconds.
4) He went to the Static Instruction Mode for 634 seconds.
5) He went to the Question Statement Page for 174 seconds.

Jeff switched 2.5 rounds between the Question Statement Page and the Static Instruction Mode (step 1-5). His program perception was positive but not very high (program perception score=3.25), which means he thought this program was only a little bit helpful for his learning. He improved slightly on his achievement post-test scores (post-test score=13).

2. Simple Animation Group

Students assigned to the Simple Animation Group had the question statement, the static instruction, and the simple animation. This study picked four cases for analysis: Ruby, Anne, Mary, and Selina. Student manipulation in the Simple Animation Group followed three basic patterns.

Pattern 2: Question Statement Page -> Static Instruction Mode -> Simple Animation Mode (students repeated going above three pages/modes)

The student repeated this order of manipulation: the Question Statement

Page, the Static Instruction Mode and then the Simple Animation Mode (e.g., Mary).

Pattern 3: Question Statement Page -> Static Instruction Mode -> Simple Animation Mode -> static instruction -> animation (students repeated going above two pages/modes)

The student went to the Question Statement Page once and then repeated switching between the Static Instruction Mode and the Simple Animation Mode (e.g., Selina and Anne).

Pattern 4: Question Statement Page -> Static Instruction Mode -> Simple Animation Mode -> Question Statement Page -> Simple Animation Mode (students repeated going above two pages/modes)

The student went through all modes once and then switched between the Question Statement Page and the Static Instruction Mode (e.g., Ruby).

Case1: Ruby (#57)- low prior knowledge student

Ruby was a student with low prior knowledge but a positive attitude towards statistics (affect score =3.00, interest score =5.75 perception of difficulty score =2.43, achievement pre-test score= 3). Ruby spent a total of 1,048 seconds on the program. The following is how she used the program:

1) She first spent 27 seconds reading the question statement.
2) She spent 338 seconds on the Static Instruction Mode.
3) She spent 487 seconds on the Simple Animation Mode.
4) She spent 9 seconds on the Question Statement Page.
5) She spent 187 seconds on the Simple Animation Mode.

Ruby started her learning by reading the question statement, then went to the Static Instruction Mode and Simple Animation Mode (step 1-3), and finally went back to the Question Statement Page and stopped in the Simple Animation Mode (step 4-5). Ruby's achievement post-test showed substantial improvement (post-test score=11). Her perception of the program was very high (program perception score=4). In other words, she considered this program very helpful.

Case2: Anne (#145)- low prior knowledge student

Anne was a student with med-low prior knowledge but a positive attitude towards statistics (affect score =4.00, interest score =5.50, perception of difficulty score =4.14, achievement pre-test score= 7). Anne spent a total of 1,510 seconds on the program. The following is how she used the program:

1) She first spent 47 seconds on the question statement.
2) She then spent 1126 seconds on the Static Instruction Mode.
3) She spent 180 seconds on the Simple Animation Mode.
4) She spent 102 seconds on the Static Instruction Mode.
5) She spent 55 seconds on the Simple Animation Mode.

Anne started her learning by reading the question statement (step 1), and then she switched 2 rounds between the Static Instruction Mode and the Simple Animation Mode (step 2-5). Anne's achievement post-test showed substantial improvement (post-test score=15). Her perception of the program was very high (program perception score=4.5). In other words, she considered this program was very helpful.

Case3: Mary (#157)- high prior knowledge student

Mary was a student with high prior knowledge and a positive attitude towards statistics (affect score =5.17, interest score =7.00, perception of difficulty score =2.29,

achievement pre-test score= 12). Mary spent a total of 870 seconds on the program. The following is how she used the program:

1) She first spent 81 seconds reading the question statement.
2) She then spent 60 seconds on the Static Instruction Mode.
3) She spent 169 seconds on the Simple Animation Mode.
4) She spent 28 seconds on the Static Instruction Mode.
5) She spent 146 seconds on the Simple Animation Mode.
6) She spent 70 seconds on the Static Instruction Mode.
7) She spent 4 seconds on the Question Statement Page.
8) She spent 1 second on the Static Instruction Mode.
9) She spent 5 seconds on the Question Statement Page.
10) She spent 233 seconds on the Static Instruction Mode.
11) She spent 73 seconds on the Simple Animation Mode.

Similar to Anne, Mary started her learning by reading the question statement (step 1), and then she went to the Static Instruction Mode and the Simple Animation Mode. After switching 2.5 rounds between the static instruction and sequence-controlled animation (step 2-6), Mary went back to the question statement. She then switched 2 rounds between the Static Instruction Mode and Simple Animation Mode (step 7-10). Mary received full points on her achievement post-test (post-test score=20). Her improvement was substantial. Her perception of the program was also very high (program perception score=4.0), which means she considered this program very helpful.

Case 4: Selina (#125)- high prior knowledge student

Selina was a student with high prior knowledge and a positive attitude towards statistics (affect score =4.67, interest score =4.50, perception of difficulty score =2.86,

achievement pre-test score= 10). Selina spent a total of 1,107 seconds on the program. The following is how she used the program:

1) She spent 76 seconds on the Question Statement Page.
2) She spent 710 seconds on the Static Instruction Mode.
3) She spent 212 seconds on the Simple Animation Mode.
4) She spent 1 second on the Static Instruction Mode.
5) She first spent 109 seconds on the Simple Animation Mode.

Selina started her learning by reading the question statement (step 1), and then switched 2 rounds between the static Simple Animation Mode and Simple Animation Mode (step 2-5). Selina's achievement post-test showed substantial improvement (post-test score=18). Her perception of the program was very high (program perception score=4). In other words, she considered this program very helpful.

3. Input Group

Students assigned to the Input Group had the question statement, the static instruction, the simple animation, and the animation allowed input manipulation. This study picked four cases for analysis: Tina, David, Ella, and Jacky. Student manipulation in the Input Group followed three basic patterns.

Pattern 5: Question Statement Page -> Static Instruction Mode-> Simple Animation Mode -> Change Inputs Mode (students repeated going above four pages/modes)

The student repeated the manipulation in this order: the Question Statement Page, the Static Instruction Mode, the Simple Animation Mode, and then the Change Inputs Mode (e.g., Tina).

Pattern 6: Question Statement Page -> Static Instruction Mode-> Simple Animation Mode (students repeated going above two pages/modes) -> Change Inputs Mode

The student went to the Question Statement Page first, and then switched between the Static Instruction Mode and the Simple Animation Mode for several rounds, and finally stayed on the Change Inputs Mode (e.g., David)..

Pattern 7: Question Statement Page -> Static Instruction Mode-> Simple Animation Mode -> Change Inputs Mode-> stay in the Static Instruction Mode or Simple Animation Mode

After going through one round of each mode, students stayed on one or two particular lower interactive mode(s) for learning (e.g., Jacky and Ella).

Case1: Tina (#74)- low prior knowledge student

Tina was a student with low prior knowledge but a positive attitude towards statistics (affect score =4.17, interest score =3.75, perception of difficulty score =4.14, achievement pre-test score= 5). Tina spent a total of 1,579 seconds on the program. The following is how she used the program:

1) She first spent 41 seconds on the Question Statement Page.
2) She spent 164 seconds on the Static Instruction Mode.
3) She spent 178 seconds on the Simple Animation Mode.
4) She spent 198 seconds on the Change Inputs Mode. She watched the animation 5 times (3 left-tailed animations, 1 two-tailed animation, and 1 right-tailed animation).
5) She spent 6 seconds on the Question Statement Page.
6) She spent 491 seconds on the Static Instruction Mode.

7) She spent 4 seconds on the Simple Animation Mode.
8) She spent 497 seconds on the Change Inputs Mode. She watched the two-tailed animation 4 times.

Tina started her learning by reading the question statement and static instruction, and then she went to the Simple Animation Mode and the Change Inputs Mode (step 1-4). She repeated the above manipulation order once (step 5-8). Tina's achievement post-test showed substantial improvement (post-test score=11). However, her perception of the program was not high (program perception score=2.5). In other words, she did not consider this program very helpful to her.

Case2: Ella (#170)- low prior knowledge student

Ella was a student with low prior knowledge but a positive attitude towards statistics (affect score =4.00, interest score =5.00, perception of difficulty score =4.57, achievement pre-test score= 7). Ella spent a total of 802 seconds on the program. The following is how she used the program:

1) She spent 58 seconds on the Question Statement Page.
2) She spent 166 seconds on the Static Instruction Mode.
3) She spent 38 seconds on the Simple Animation Mode.
4) She spent 58 seconds on the Change Inputs Mode. She watched the two-tailed animation twice.
5) She spent 31 seconds on the Question Statement Page.
6) She spent 423 seconds on the Static Instruction Mode.
7) She spent 24 seconds on the Question Statement Page.

Ella started her learning by reading the question statement and static instruction, and then she went to the Simple Animation Mode and the variable -controlled Simple Animation Mode (step 1-4). Then she switched between the Question Statement Page and

the Static Instruction Mode, and stayed on the Static Instruction Mode for a very long time (step 5-7). Ella's achievement post-test showed minor improvement (post-test score=10). Her perception of the program was highly positive (program perception score=4.25). In other words, she considered this program very helpful to her.

Case3: David (#118)- high prior knowledge student

David was a student with low prior knowledge and a positive attitude towards statistics (affect score =5.00, interest score =3.00, perception of difficulty score =3.71, achievement pre-test score= 12). David spent a total of 1,155 seconds on the program. The following is how he used the program:

1) He spent 54 seconds on the Question Statement Page.
2) He spent 723 seconds on the Static Instruction Mode.
3) He spent 164 seconds on the Simple Animation Mode.
4) He spent 82 seconds on the Static Instruction Mode.
5) He spent 66 seconds on the Simple Animation Mode.
6) He spent 66 seconds on the Change Inputs Mode. He watched the left-tailed animation once.

David started his learning by reading the question statement (step1), and then he switched twice between the Static Instruction Mode and the Simple Animation Mode (step 2-5). He stopped the learning activity by staying on the variable controlled mode (step 6). David's achievement post-test showed substantial improvement (post-test score=19). His perception of the program was positive (program perception score=3.25). In other words, he considered this program helpful.

Case4: Jacky (#70)- high prior knowledge student

Jacky was a student with high prior knowledge and a positive attitude towards statistics (affect score =5.00, interest score =4.75, perception of difficulty score =4.71,

pre-test score= 10). Jacky spent a total of 809 seconds on the program. The following is how he used the program:

1) He spent 38 seconds on the Question Statement Page.
2) He spent 545 seconds on the Static Instruction Mode.
3) He spent 21 seconds on the Simple Animation Mode.
4) He spent 7 seconds on the Change Inputs Mode. He watched the left-tailed animation once.
5) He spent 16 seconds on the Question Statement Page.
6) He spent 128 seconds on the Simple Animation Mode.
7) He spent 1 second on the Question Statement Page.
8) He spent 53 seconds on the Static Instruction Mode.
9) He spent 7 seconds on the Simple Animation Mode.

Jacky started learning by reading the question statement, and then he went to the Static Instruction Mode, Simple Animation Mode and Change Inputs Mode(step 1-4). Then he switched among the Question Statement Page, the Static Instruction Mode and the Simple Animation Mode (step 5-9). Jacky's achievement post-test showed only minor improvement (post-test score=11). His perception of the program was not high (program perception score=2.75). In other words, he did not consider this program very helpful to him.

4. Practice Group

Students assigned to the Practice Group had the question statement, the static instruction, the simple animation, animation allowed input and practice manipulation. This study picked five cases for analysis: Maggie, Joy, Bill, and Michelle. Student manipulation in the Practice Group followed four basic patterns.

Pattern 8: Question Statement Page -> Static Instruction Mode -> Simple Animation Mode -> Change Inputs Mode (students repeated going above four pages/modes) -> Practice Mode

The student repeated the manipulation in the order: the Question Statement Page, the Static Instruction Mode, the Simple Animation Mode, the Change Inputs Mode, and then the Practice Mode (e.g., Maggie and Joy).

Pattern 9: Question Statement Page -> Static Instruction Mode -> Simple Animation Mode -> Change Inputs Mode-> practice (students repeated going above two pages/modes)

After a quick look at each button, the student switched between the Change Inputs Mode and the Practice Mode for learning (e.g., Bill).

Pattern 10: Question Statement Page -> Static Instruction Mode -> Simple Animation Mode -> Change Inputs Mode-> practice -> static instruction or Simple Animation Mode (students repeated going above two pages/modes)

After a quick look at each mode, the student stayed in the static Simple Animation Mode or the Simple Animation Mode for learning (e.g., Michelle).

Pattern 11: Question Statement Page -> Static Instruction Mode -> Simple Animation Mode

The students did not go through all modes but only stayed a few seconds in some of the lower level modes (e.g., Peggy).

Case1: Maggie (#135)- low prior knowledge student

Maggie was a student with very low prior knowledge but a positive attitude towards statistics (affect score =3.33, interest score =3.75, perception of difficulty score =2.86, achievement pre-test score= 4). Maggie spent a total of 2,458 seconds on the program. The following is how she used the program:

1) She spent 100 seconds on the Question Statement Page.
2) She spent 1261 seconds on the Static Instruction Mode.
3) She spent 365 seconds on the Simple Animation Mode.
4) She spent 8 seconds on the Question Statement Page.
5) She spent 136 seconds on the Static Instruction Mode.
6) She spent 89 seconds on the Simple Animation Mode.
7) She spent 260 seconds on the Change Inputs Mode. She watched the left-tailed animation once, the right-tailed animation once, and the two-tailed animation twice.
8) She spent 239 seconds on the Practice Mode. She did 3 practices (1 left-tailed, 1 right-tailed and 1 two-tailed practice) and got 2 of them correct.

Maggie started her learning by reading the question statement, and then she went to the static instruction and sequence-controlled animation (step 1-3). Then she repeated this manipulation once (step 4-6). Finally she went to the Change Inputs Mode (step 7) and then did 3 practices in the Practice Mode (step 8). In her first practice set, she got all answers correct on her third try. In her second practice, she tried to input the answers 10 times but did not get them all correct. In the third practice, she got all the answers correct on her second try. Maggie‘s achievement post-test showed substantial improvement (post-test score=16). Her perception of the program was highly positive (program perception score=4). In other words, she considered this program very helpful to her.

Case2: Joy (#87)- low prior knowledge student

Joy was a student with very low prior knowledge but a positive attitude towards statistics (affect score =3.67, interest score =5.25, perception of difficulty score =2.71, achievement pre-test score= 7). Joy spent a total of 1,239 seconds on the program. The following is how she used the program:

1) She spent 54 seconds on the Question Statement Page.
2) She spent 64 seconds on the Static Instruction Mode.
3) She spent 7 seconds on the Simple Animation Mode.
4) She spent 2 seconds on the Change Inputs Mode. She watched the left-tailed animation once
5) She spent 1 second on the Simple Animation Mode.
6) She spent 2 seconds on the Practice Mode for 1 left-tailed question. She did not input any answers.
7) She spent 1 second on the Simple Animation Mode.
8) She spent 3 seconds on the Static Instruction Mode.
9) She spent 22 seconds on the Question Statement Page.
10) She spent 335 seconds on the Static Instruction Mode.
11) She spent 49 seconds on the Simple Animation Mode.
12) She spent 146 seconds on the Change Inputs Mode. She watched the left-tailed animation three times.
13) She spent 33 seconds on the Static Instruction Mode.
14) She spent 26 seconds on the Change Inputs Mode. She watched the left-tailed animation once.
15) She spent 48 seconds on the Static Instruction Mode.
16) She spent 34 seconds on the Change Inputs Mode. She watched the right-tailed animation twice.

17) She spent 92 seconds on the Practice Mode. She did 2 left-tailed practices and did not get all answers correct in her 10 trials.
18) She spent 8 seconds on the Static Instruction Mode.
19) She spent 146 seconds on the Simple Animation Mode.
20) She spent 17 seconds on the Practice Mode. She did 1 right-tailed practice but did not input any answers.
21) She spent 1 second on the Question Statement Page.
22) She spent 148 seconds on the Practice Mode. She did 3 right-tailed practices. In her first practice, she got all answers correct in her first trial; in her second practice, she got all answers in her second trial. She did not input answers in her third practice.

Joy first took a quick look at each mode (step 1-4) and then stayed in the Static Instruction Mode for a long time (step 10). Then she went to the Simple Animation Mode and the Change Inputs Mode (step 11-12). She switched 2 rounds between the static instruction and the Change Inputs Mode (step 13-16). Then she started to do the practice, but she did not get the answer correct (step 17). As a result, she went back to the static instruction and stayed in the Simple Animation Mode for a long time (step 19). Then she did her second practice (step 20) and third practice (step 22). Joy successfully solved the questions in two of the practices (step 22). Joy's achievement post-test showed substantial improvement (post-test score=18). Her perception of the program was highly positive (program perception score=4.25). In other words, she considered this program very helpful to her.

Case 3: Bill (#71)- high prior knowledge student

Bill was a student with high prior knowledge and a positive attitude towards statistics (affect score =5.83, interest score =5.50, perception of difficulty score =4.14,

achievement pre-test score= 13). Bill spent a total of 2,223 seconds on the program. The following is how he used the program:

1) He spent 87 seconds on the Question Statement Page.
2) He spent 304 seconds on the Static Instruction Mode.
3) He spent 285 seconds on the Simple Animation Mode.
4) He spent 163 seconds on the Change Inputs Mode. He watched 4 left-tailed animations.
5) He spent 105 seconds on the Practice Mode for 1 left-tailed question and 2 right-tailed questions. He tried to input answers twice in his first practice but did not get all of them correct. He tried to input answers six times in his second practice but still did not get all of them correct. He did not input any answers in his third practice.
6) He spent 1 second on the Simple Animation Mode.
7) He spent 59 seconds on the Change Inputs Mode. He watched 1 right-tailed animation and 2 left-tailed animations.
8) He spent 48 seconds on the Practice Mode for 2 right-tailed questions. He tried to input answers six times in his first practice but did not get all of them correct. He did not input any answers in his third practice.
9) He spent 37 seconds on the Change Inputs Mode. He watched 1 right-tailed animation.
10) He spent 1 second on the Simple Animation Mode.
11) He spent 43 seconds on the Change Inputs Mode. He watched 1 right-tailed animation.
12) He spent 133 seconds on the Practice Mode for 3 right-tailed questions. In his first practice, he got all answers correct in his third trial; in his second practice, he got all answers in his first trial. He did not input answers in his third practice.

13) He spent 44 seconds on the Change Inputs Mode. He watched 2 right-tailed animations.
14) He spent 46 seconds on the Practice Mode for 2 right-tailed questions. He tried to input answers three times in his first practice but did not get all them correct. He did not input any answers in his third practice.
15) He spent 51 seconds on the Change Inputs Mode. He watched 2 right-tailed animations.
16) He spent 346 seconds on the Practice Mode for 2 right-tailed questions, 1 two-tailed question, and 2 left-tailed questions. In his first practice, he got all answers correct in his fourth trial; in his second practice, he got all answers in his second trial; in his third practice, he got all answers correct in his first trial; in his fourth practice, he got all answers in his second trial. He did not input answers in his fifth practice.
17) He spent 108 seconds on the Change Inputs Mode. He watched 1 left-tailed animation.
18) He spent 4 seconds on the Simple Animation Mode.
19) He spent 22 seconds on the Static Instruction Mode.
20) He spent 41 seconds on the Change Inputs Mode. He watched 2 left-tailed animations.
21) He spent 295 seconds on the Practice Mode for 5 left-tailed questions. In his first and second practice, he got all answers correct in his second trial. In his third practice, he got all answers correct in his first trial; in his fourth practice, he got all answers correct in his second trial. He did not input answer in his fifth practice.

Bill first went through each mode quickly (step1-4). Then he did one practice but did not get all answers correct. Then he switched four rounds between the Change Inputs Mode and the Practice Mode (step 7-8; step 9-12; step 13-14; step 15-16; step 17-21). He

got all the answers correct in his last two practices. He sometimes clicked and stayed on the static instruction and sequence-Simple Animation Mode, but he stayed on these two modes for very short spans of time. It seems that the Change Inputs Mode helped him the most. Bill's achievement post-test showed only minor improvement (post-test score=15). His perception of the program was highly positive (program perception score=4). In other words, he considered this program very helpful to him.

Case 4: Michelle (#103): High prior knowledge and high engagement

Michelle was a student with very high prior knowledge and a positive attitude towards statistics (affect score =4.50, interest score =5.75, perception of difficulty score =3.14, achievement pre-test score=12). Michelle spent a total of 1,258 seconds on the program. The following is how she used the program:

1) She spent 31 seconds on the question statement.
2) She spent 398 seconds on the static instruction.
3) She spent 74 seconds on the Simple Animation Mode.
4) She spent 245 seconds on the Change Inputs Mode. She watched 2 right-tailed animations.
5) She spent 246 seconds on the Practice Mode. She did 1 two-tailed practice and 2 right-tailed practices. In her first practice, she got all answers correct in her first trial; in her second practice, she got all answers in her second trial. She did not input answers in her third practice.
6) She spent 28 seconds on the static instruction.
7) She spent 236 seconds on the Practice Mode. She did 1 left-tailed practice, 1 right-tailed practice, and 1 two-tailed practice. She got all answers correct in the first trial in these 3 practices.

Michelle started her learning by reading the question statement, and then she went to the static instruction, sequence-controlled animation and variable-controlled animation (step 1-4). Then she did three practices in the Practice Mode, and she got all answers correct in two of the practices (step 5). Then she went back to the static instruction (step 6). She did another three practices and got all of them correct (step 7). Michelle's achievement post-test showed substantial improvement (post-test score=16). Her perception of the program was highly positive (program perception score=4.25). In other words, she considered this program very helpful to her.

Case5: Peggy (#155)- high prior knowledge student

Peggy was a student with high prior knowledge but a negative attitude towards statistics (affect score =3.67, interest score =2.75, perception of difficulty score =4.29, pre-test score= 10). Peggy spent a total of 977 seconds on the program. The following is how she used the program:

1) She spent 79 seconds on the Question Statement Page.
2) She spent 541 seconds on the Static Instruction Mode.
3) She spent 357 seconds on the Simple Animation Mode.

Peggy started her learning by reading the question statement, and then she went to the Static Instruction Mode and Simple Animation Mode (step 1-3). She stayed in the lower interactive modes for learning; she did not go to the Change Inputs Mode or Practice Mode. Peggy improved a little bit on her achievement post-test (post-test score=11). Her perception of the program was positive but not too high (program perception score=3.75). In other words, she considered this program somewhat helpful.

Summary

The above section describes 17 students' detailed manipulation of the program and generates 11 manipulation patterns. These manipulation patterns could be characterized and summarized in five categories to better describe student manipulation:

1. Category 1: Switching between the static material and the dynamic material (Pattern 3 in the Simple Animation Group).
2. Category 2: Using the lower interactive modes first and then the higher interactive ones (Pattern 6 in the Simple Animation Group and Pattern 8 in the Practice Group).
3. Category 3: Going through each mode quickly and then staying in higher interactive modes for learning (Pattern 4 in the Simple Animation Group and Pattern 9 in the Practice Group).
4. Category 4: Going through each mode quickly and then staying in the lower interactive modes for learning (Pattern 6 in the Simple Animation Group, Pattern 7 in the Input Group and Pattern 10 in the Practice Group).
5. Category 5: Repeating by the order of the buttons (Pattern 1 in the Static Group, Pattern 2 in the Simple Animation Group and Pattern 5 in the Input Group).
6. Category 6: Quickly picking a mode for learning but not going through all modes (Pattern 11 in the Practice Group).

This study anticipated that the best cases were those who chose one mode or one particular order of modes according to their preference. However, some students clicked on the modes just by meaninglessly following the button order which was shown on the program interface (e.g., category 5). In addition, some students just went through the program very quickly without thoroughly watching animations (e.g., category 4). Some even skipped the practice opportunities (e.g., category 6). The Web log data shows that not all students manipulate the program as expected.

SUMMARY

The hypothesis testing results showed that the increasing animation interactivity enhanced the intermediate-level learning. Animation interactivity impacted understanding (p=.006) and lower-level applying (p=.042). Animation interactivity did not impact student confidence or program perception. The regression analysis indicated that student prior knowledge and interest were the most important predictors of students' achievement post-test scores instead of animation manipulation. The regression tree showed that there were interactions among student interest, prior knowledge, and animation manipulation on the achievement post-test scores. The case analysis showed that not all students manipulated the interactive animation program as expected (e.g., students in the Practice Group did not thoroughly take advantage of the Practice Mode). This could be one reason why the Practice Group did not outperform other groups in the achievement post-test. The following chapter discusses and connects these findings in detail.

Chapter 5 Discussion and Conclusion

This chapter includes a summary of the study, and a discussion of the results, implications, limitations, and conclusions.

SUMMARY OF THE STUDY

The study was designed to explore the impact of animation interactivity on undergraduate student learning in introductory statistics. It was a one-way design where the independent variable was animation interactivity. In addition to a control group (Static Group) provided with only static materials, there were three groups with different levels of animation interactivity: 1) Animation without interactivity (Simple Animation Group), 2) animation with input manipulation (Input Group), and 3) animation with practice and feedback (Practice Group). A sample of 123 college students participated in the study and was randomly assigned into groups. They gathered in a computer lab to work with the interactive animation program and took online surveys and tests for evaluation. Students were expected to learn *Principles of Hypothesis Testing* (concepts of type I error, type II error and *p*-value). The data collected in this study included student learning attitude, achievement and confidence pre-test scores, achievement and confidence post-test scores, and program perception. Also, student manipulation of the program was recorded as Web log data. The data was analyzed using the following techniques: Multivariate analysis (MANOVA), univariate analysis (ANOVA), regression analysis, regression tree analysis, and case analysis. The findings are as follows: 1) Student understanding improvement (p=.006) and lower-level applying improvement (p=.042) were impacted by animation interactivity, 2) student confidence and program perception were not impacted by animation interactivity, 3) the regression analysis indicated that student prior knowledge and interest were the most important predictors of

students' achievement post-test scores rather than animation manipulation, and 4) the regression tree illustrated that there were interactions among student interest, prior knowledge, and animation manipulation on the achievement post-test scores. The case analysis showed that not all students manipulate the interactive animation program as expected, and this could be one reason why the Practice Group did not outperform other groups on the achievement improvement. The following section discusses the findings in detail.

DISCUSSION OF THE RESULTS

Question1: Animation interactivity and achievement improvement

Is there any difference in achievement improvement among students who use different interactive levels of an animation program?

The first research question was concerned with animation interactivity and student achievement improvement. The testing of hypotheses 1 through 4 provided answers for this research question. The following sections discuss the findings from these tests.

The interactiviy effect existed on the intermediate-level learning

The testing results showed that the increase of animation interactivity did enhance student achievement improvement on student understanding and lower-level applying, but not on remembering and higher-applying. In other words, the increase of animation interactivity impacted student learning on the intermediate-level learning (Figure 5.1).

When students were asked to perform simple tasks related to remembering or recalling, static materials were just as effective as interactive animations. But with the increase of the level of knowledge delivered (e.g., asking students to interpret terminology definitions or do calculations based on principles learned), the animation interactivity showed higher beneficial effects. This effect went away when the knowledge

delivered became too complicated and abstract (e.g., asking students to interpret a solution or answer). This result was consistent with prior research that has found that no differences existed between animation and static graphics when considering low-level recall information (Rieber, 1990). However, it showed limitations of prior statements of animation's beneficial effects (Rieber, 1990; Mayer & Anderson, 1992) by arguing that increasing animation interactivity did not impact all levels of learning.

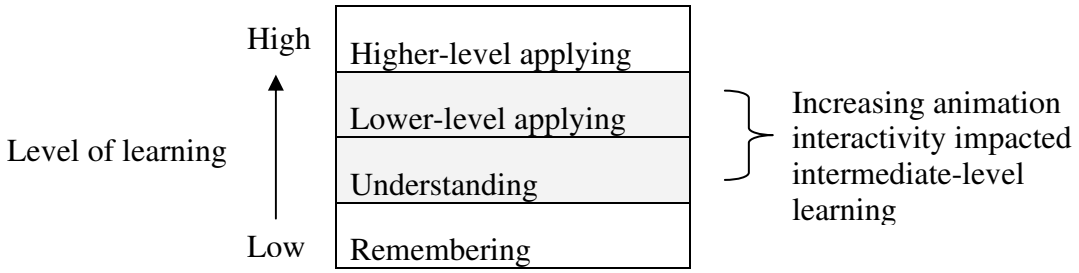

Figure 5.1 Animation interactivity and knowledge levels

The results expanded Evans & Gibbons' interactivity effect theory. The interactivity effect theory suggests that higher interactivity makes for better learning. However, this study showed that the interactivity effect did not necessarily exist on all levels of learning. It was possible that in terms of the lowest-level of learning (e.g. remembering), students can do it well without the provision of interactive animation. And for the highest-level of learning (e.g. higher-level applying), providing only interactive animation was not enough to cultivate student learning. Students may need other supports such as accompanied instructional strategies for better learning.

In addition, this study explained why some researchers have argued that interactive multimedia is helpful (Rieber, 1990; Mayer, 2009) while some have not (Boucheix & Schneider, 2008). This might be because these researchers targeted at

different levels (or types) of learning or unclearly defined the knowledge delivered. Consequently, they produced contradictory results and made limited claims. For example, Mayer (2009) argued that the use of interactive animation helped students with learning transfer. However, there were limitations to this argument. The concept of transfer could be divided into two categories: near transfer and far transfer, corresponding to "lower-level applying" and "higher-level applying" in this study. When the researchers claimed that learning transfer was improved with the use of interactive animation, which part of transfer did the study mean was improved? Because measurement was not well defined in the prior study (Mayer, 2009), we cannot answer this question. But in this study, evidence indicated that it was near transfer (lower-level applying) that was improved by the use of interactive animation. The study expanded upon prior research by dividing learning into several levels, and showed the results of animation interactivity at different levels of learning.

Higher interactivity did not necessarily optimize achievement improvement

This study originally assumed that higher animation interactivity would improve student learning based on the interactivity effect theory (Evans & Gibbons, 2007). Based on the assumption, students in the Practice Group should have had the best learning improvement since they were provided with the highest interactive mechanism (animation with practice and feedback). However, the results did not support it. The post hoc tests showed that 1) student understanding improvement in the Animation, Inputs and Practice Group was significantly better than that in the Static Group, but 2) there was not any difference among the three treatment groups. The first finding indicated that animation enhanced student understanding more than the static instruction. The second finding indicated that high interactivity did not always help learning. The second finding was an interesting one which did not meet my prediction.

To better understand why the highest interactivity (animation with practice and feedback) did not optimize student learning, students' Web log data were examined carefully. This study picked five students randomly from the Practice Group for analysis and found that some students did not manipulate the animation program as expected. The best case was that the student went through all modes and stayed in the higher interactive modes for learning, but not all students did this. For example, some students did take advantage of the Practice Mode to get feedback (e.g., category 3) but some just stayed in the lower interactive level modes and skipped the higher interactive modes (e.g., category 4 and category 6). Only a small group of students made full use of the program. Students' unexpected manipulation made the provision of the high interactive modes ineffective. This could be why students in the Practice Group did not outperform students in other groups in the achievement post-test. If students in the Practice Group could engage the program as expected, they should have higher post-test achievement scores and improvement.

The reasons for students' unexpected manipulation of animation are complicated (Boucheix et al., 2008). Students' shortage of cognitive and information management skills, or a lack of computer confidence and competence may have prevented them from successfully using the animation program. In this study, the Practice Mode was built with several sets of drop-down menus, input boxes and question-and-answer trials, which might have overwhelmed some students and made them unwilling to interact. Since the Practice Group was given a program interface much more complex than the other groups received, this study did not exclude the possibility that some of the students were not capable of taking advantage of the highest level of interactivity. In addition, the feedback given in the Practice Mode might have limited students' learning improvement. The feedback given in the Practice Mode was a type of "try-again" feedback which informed

learners whether their answers were correct (Mory, 2004). This type of feedback did not provide elaboration or explanation to students on why their answers were right or wrong. Consequently, in this study, though students knew that they made a mistake in a question, they did not know why they got it wrong. They needed to go through the materials and animations again to find a correct answer by themselves; however, some students might not be capable of doing this. And some even lacked interest and motivation since it was a lab-based experiment irrelevant to their school learning.

In contrast, the Change Inputs Mode provided students with simple and controllable animation interaction. Students inputted different variables and the program generated animations. They were not stuck in the question-and-answer practices. For most students, the Change Inputs Mode served as a "simple verification feedback and knowledge result" (Mory, 2004). The simple interaction made learning easy and therefore students showed pretty good improvement.

The time limitation could also have factored into why the students in the Practice Group did not perform well. This study was a lab-based experiment with two-hour time period. It was very tight for students in the Practice Group to go through different modes and have enough time to practice and achieve learning. If the time for the experiment is expanded to days or weeks, it is possible that the Practice Group will have better learning outcomes.

Question2: Animation interactivity and confidence improvement

Is there any difference in confidence improvement among students who use different interactive levels of an animation program?

The second research question was concerned with animation interactivity and student confidence improvement. The testing of hypotheses 5 through 8 provided answers for this research question. The results indicated that there were no significant differences

in confidence improvement among students who used different interactive levels of the animation program. In other words, increasing animation interactivity did not necessarily enhance student confidence. This finding differed from previous studies which stated that students felt more confident with advice on control selections because it helped them make better choices (Litchfield, 1993).

Confidence, or self-efficacy, is defined as the belief in one's capabilities to complete a task. It is a predictor for how one will execute the courses of actions required to manage prospective situations (Bandura, 1997). Bandura argued that providing a pedagogic strategy such as inquiry-based learning or project-based learning fosters student self-efficacy, and he also claims that mastery experience is the most robust source of gaining confidence. In this study, the Practice Mode provided to the Practice Group was intended to create an environment where the students could gain mastery through repeated practices. The best case was that after students read the question statement and static instruction, they switched between animations and conducted practices until they had mastered the concepts. However, not all students in the Practice Group manipulated this program as expected. As mentioned above, the Web log data showed that some students just went through the program very quickly without thoroughly watching the animations; some even skipped the practice opportunities. The Web log data provided the following explanations for why student confidence improvement was not enhanced by the increase in interactivity.

Complex interface decreased student confidence

Increasing the animation interactivity made the program interface more complicated. Compared with students in the Static, Simple Animation and Input Groups, students in the Practice Group needed more cognitive skills to process the information delivered from the program. Students who were not capable of manipulating the program

might have been intimidated by the complex interface with buttons, text boxes and drop down menus. For example, Peggy in the Practice Group only watched two of the four modes, and then quit the program quickly. Obviously she was not interested in using the program to learn new information. She clicked on modes that were simple for her and then exited. One possible reason for her quick quit was that the complex interface might have frightened her and reduced her willingness to learn, thus limiting her confidence improvement. The other possible reason was that this was an experiment conducted in a laboratory setting rather than a real-world classroom, and she did not have motivation to spend time exploring the program and improving confidence.

The unsuccessful practice trials decreased student confidence

The Practice Mode was designed to provide students with opportunities for repeated practices. However, this design might have negatively influenced student confidence in learning statistics. Those students who easily got all answers correct in their first practices probably had dramatic improvement in their confidence. In contrast, those students who did poorly in their first couple of practices might have confidence decreased. Some students even gave up practicing when they continued to get incorrect messages. For example, Maggie engaged very much in using the program for learning statistics. However, for some reason, she did not get all the answers correct in her last practice of a two-tailed test question. After a long time trial, she decided to give up practicing and went to the post-test. Her unsuccessful trials in the practices very possibly frustrated her and decreased her confidence improvement.

Though researchers have agreed that the increase of program interactivity enhances learning, the results of question 1 and question 2 showed that the program complexity, students' cognitive ability for processing information, and the design of practice procedure mediate learning. Educators should consider these factors carefully

when integrating interactive animation in educational settings. For example, while building a practice mechanism into a program, the designers are urged to make the interface simple, prepare pre-training sessions, provide real-time feedback with sufficient explanations, and give appropriate encouragement to engage students to learn.

Question3: Animation interactivity and program perception

Is there any difference in animation perception among students who use different interactive levels of an animation program?

The third research question was concerned with animation interactivity and student program perception. The testing of hypothesis 9 provided answers for this research question. The results indicated that there were no significant differences in program perception among students who used different interactive levels of an animation program (p=.352). In other words, the increase of animation interactivity did not enhance student program perception. The perception survey showed that the average perception scores for the three experimental groups were all above 3.40 (range: 1-5), which means that students had a rather positive perception of interactive animation. But the difference among groups was not very clear.

The insignificance of program perception among groups might be due to different reasons. Because this was a lab-based experiment lasting only two hours, time may have been an important factor. The time might have been too short for students to adequately experience the animation and learn the particular unit. Consequently, the effect of the animation interactivity was not clear. Another factor might have been participant selection. Participants were randomly drawn from the subject pool, and all tasks in this study were not relevant to their school learning. As a result, some students might not have had adequate motivation to use the animation program seriously and thus their perception difference was limited.

Web log data analysis

The interaction of prior knowledge and animation manipulation on learning achievement

Although looking at the individual differences and student animation manipulation were not a major purpose of this study, the Web log data analysis was used to provide additional information for hypothesis testing results. In the regression analysis, student prior knowledge and interest were found to be the best predictors of their post-test achievement scores. The tree analysis supported these results by showing student prior knowledge and interest as the two most important splitters in student post-test performance. Moreover, it indicated that the provision of animation only impacted students with low interest in statistics (Figure 4.1).

Watching animation did not enhance learning for students with high interest in statistics, but it did for students with low interest. More than 76% of students with high interest and high prior knowledge achieved high post-test scores, which was above average (Node 6 in Figure 4.1). That is, most students with high interest and high prior knowledge had high post-test performance no matter how they used the animation. However, for students who had low interest, high frequency of watching animation program appeared to help achieve high post-test scores. Students with low interest and low frequency of watching animation had only a 28.6% rate of high post-test scores (Node 3 in Figure 4.1), while students with low interest and high frequency of watching animation achieved a 66.7% rate of high post-test scores (Node 4 in Figure 4.1). The numbers showed that watching interactive animation might not help high-interest students but it greatly helped low-interest students to achieve high post-test scores.

In addition, the tree analysis also showed that students with low interest outperformed students with high interest if they had high frequency of watching an animation program. Students with low interest and high frequency of watching animation achieved a 66.7% rate of high post-test scores (Node 4 in Figure 4.1), while students with high interest but low prior knowledge had a rate of only 43.8% (Node 5 in Figure 4.1). In other words, low-interest students had the possibility to outperform high-interest students if they were provided with an animation.

The decision tree analysis generated three important predictors (splitters) to students' post-test scores: 1) Interest, 2) prior knowledge, and 3) frequency of watching animation. These three were the most important variables among all while integrating an animation into a statistics classroom. Students with high learning interest (especially those also with high prior knowledge) had a very high chance to achieve high post-test scores. Providing an animation to this group of students might be helpful but the effect could be very marginal. In contrast, for the low-interest students, increasing their frequency of watching an animation greatly enhanced their chance of achieving high post-test scores. The provision of animation made a low-interest student's learning totally different. This result indicated that instructors should be aware of student learning interest before integrating an animation into the classroom. Providing the opportunity of watching animations might be very helpful for low-interest learners since the animation could be a motivator to them. But the animation might be redundant if learners themselves have high interest and are capable of learning without supports.

The results showed both the limitations and affordances of interactive animation. Animation did improve learning but not for all students over all conditions. Instructors need to factor in student individual differences to optimize the use of animation

programs. Integrating animation into learning without carefully considering individual differences will marginalize the benefits.

IMPLICATIONS

The current study provided theoretical, methodological and pedagogical implications for the use of interactive animation.

Theoretical implications

With regard to theoretical implications, this study drew on research in interactive animation to explore the interactivity effect. By examining the relationship between animation interactivity, animation manipulation and individual differences (learning interet and prior knowledge), this study offered the following theoretical implications:

1. Expanding Mayer's multimedia theory

This study expanded Mayer's multimedia theory by considering the interactivity effect of an animation program. Mayer's experiments and arguments mostly focused on the effect of static media (e.g., static diagrams and printed texts). Interactive multimedia (e.g., interactive visualization or animation) was involved in some of his studies; however, they were not discussed at length. This study focused on the use of an interactive animation program and the effect of interactivity, which not only supplemented Mayer's multimedia theory but opened a new window for multimedia research on the exploration of interactive animation, cognitive levels and individual differences.

2. Expanding the interactivity effect theory

There is one assumption underlying all interactivity theories: the higher the interactive level, the better the product (Evans & Gibbons, 2007). However, this study found that there were factors influencing learning when animation interactivity was

presented. Program interactivity, prior knowledge, interest and cognitive ability are all possible factors relevant to the efficiency of animation use. These factors play key roles in the success of animation use.

3. Providing a possible explanation for contradictions in previous studies

Prior studies of animation have had inconsistent findings: Some researchers have argued that user-control functions were helpful to learning (Rieber, 1990) while others have not (Boucheix & Schneider, 2008). This study provided a possible explanation for these contradictions. This study argued that research inconsistencies might have occurred because researchers targeted different levels or types of learning, which consequently resulted in different outcomes. In short, the use of interactive animation did not help students learn at all levels of knowledge, and it was not always helpful for all students and in all conditions.

Methodological implications

Most prior research has explored the use effectiveness of interactive multimedia using only ANOVA/ MANOVA approaches (Mayer, 2009). The present study examined the students' use of interactive animation from a different perspective. In addition to using the traditional ANOVA/MANOVA techniques, this study used the regression tree and case analysis technique to analyze Web log data. The findings from these techniques were supplementary data to better explain students' online behaviors. In the practice of many multimedia studies, Web log analysis is encouraged for better understanding of how students manipulate online tools. This research is a sample of integrating innovative Web log data analysis with the traditional data analysis. More studies need to be done to help the integration of these analyses mature.

Pedagogical implications

The findings of this study gave rise to several pedagogical implications for the use of an interactive animation program in a statistics classroom. To optimize the use of an interactive animation program, teachers and educational prationers are expected to accomplish the following tasks:

1. Evaluate the necessity of integrating an interactive animation program.

Teachers must evaluate student needs for using an interactive animation. Providing interactive animation may not always be helpful for all students. For example, for those who have very high interest and prior knowledge, the helpfulness of an interactive animation may be limited. Teachers need to consider other approaches or accompany animation with advanced instructional strategies to help students learn. Also, given the cost of designing animnations compared to traditional static materials, teachers have to make sure that the use of an animation is necessary in the classroom.

2. Evaluate students' cognitive and information management skills before adopting an interactive animation program.

Teachers must consider students' cognitive and information management skills when assessing their ability to use an interactive animation program. If students do not have the ability to deal with a complex program interface, a high interactivity program may discourage and frustrate students. Teachers and program designers need to find a balance point between students' capability and animation complexity when choosing or designing an interactive animation program.

3. Prepare students to effectively manipulate an interactive animation program.

Teachers are encouraged to provide manuals, guides, or instructions before letting students use the animation program. This preparation can make students familiar with the

program interface quickly and decrease their confusion and frustration when using the animation program.

4. Adopt instructional strategies to optimize the effectiveness of an interactive animation program.

To enhance students' interactions with the animation program, teachers are encouraged to use a set of strategies that accompany the use of an interactive animation program. For example, collaboration strategies can be included so that students can work with the animation program in groups. Inquiry-based learning can also be used to guide students' interactions with the animation program. Such strategies can improve the effectiveness of animation and therefore increase learning outcomes.

LIMITATIONS

Several limitations need to be acknowledged in the interpretation of these results. First of all, this research was conducted in a laboratory setting rather than in a real classroom situation, thus students may have had different motivations and may have exhibited different behaviors than those experienced during an actual statistics class. Second, this research was restricted to one unit from an introductory statistics course. Students may have had different responses if this particular unit had been exchanged for another one. Third, this particular animation program was established with a limited scope and relitively simple functions, and this may have reduced the interactivity effect . Fourth, some of the instruments (such as the achievement and confidence tests) were not perfectly reliable (lower than .70 or higher than .90), and this could be an issue for the results of the study. Finally, the study focused on a fairly small number of participants (about 30 students per group), so that the hypothesis testing results might not be very stable. Due to these limitations, the generalization of this study must be conservative.

CONCLUSIONS

Based on the above discussion, this study offers the following conclusions. First of all, increasing animation interactivity appeared to enhance student achievement improvement at the intermediate level of learning (understanding and lower-level applying). Secondly, student confidence improvement and animation perception were not impacted by the increase of animation interactivity. Thirdly, student prior knowledge and interest towards statistics were the most important predictors of students' post-test scores rather than animation manipulation. Finally, there were interactions among student individual differences (interest and prior knowledge) and animation manipulation on the learning outcomes.

This study not only revealed the limitations of interactive animation, but also showed the necessity for instructors to improve the effectiveness of interactive animation programs. The first necessity is that instructors or program designers have to find a balance between animation interactivity and program complexity. High interactivity is a common goal for animation designers, but it does not always improve student learning. Increased interactivity increases program complexity, which may decrease learners' ability in using the program. Animation designers must be very careful when adding interactive functions to avoid negatively affecting learner motivations. The other necessity is that instructors have to be aware of learners' individual differences while integrating an animation. An animation program may have different effects due to students' different interest and prior knowledge. Based on the above discussion, further research is suggested to explore the following directions:

1. Explore students' balance between animation interactivity and program complexity.

2. Explore the interactivity effect in a real-world classroom with instructional strategies instead of in a laboratory.
3. Explore the impact of the length of time of the interaction with interactive animation on learning.
4. Explore the impact of student prior knowledge or individual differences on their use of interactive animation.
5. Revise current instruments and adding more constructs (e.g., learning or computer preference) into this study.
6. Expand the number of subjects for better statistical power.
7. Conduct qualitative research studies to explore students' thoughts concerning the interactive animation program.

Last, given the cost of designing interactive animations compared to traditional materials, one question an instructor designer will ask is: "Do I really need to use interactive animation?" It is important to point out again that animation should be used only when learners have cognitive necessity and when the animation is carefully designed and integrated. Here I want to express the same thoughts as Tversky, Morrison and Betrancourt's argument by quoting from their paper in 2002:

> Participants can study those aspects of the animation that they need without suffering through portions they already understand. Then, carefully crafted animations can be apprehended, those that highlight the discrete and high-level steps and those that depict the analog and microsteps that animations seem well-suited to portray and convey. Like all good things, animation must be used with care. (p.258)

Appendix A – The pilot study results

Table 1 Paired T-tests for student achievement improvement

	Pre-test		Post-test		Paired T-test	
	Mean	Sd.	Mean	Sd.	T	Sig.
Remembering-/understanding-level knowledge	2.72	.57	2.67	.49	.0369	.717
Applying-level knowledge	4.06	2.21	5.17	1.50	-2.938	.009 **

*Maximum possible score on remembering-/understand-leveling knowledge is 3

*Maximum possible score on applying-level knowledge is 7

Table 2 Paired T-tests for student achievement improvement

	Pre-test		Post-test		Paired T-test	
	Mean	Sd.	Mean	Sd.	T	Sig.
Remembering-/understanding-level knowledge	3.75	.28	3.89	.26	-.167	.110
Applying-level knowledge	2.80	.82	3.26	.55	-2.57	.020**

*Maximum possible score on student confidence is 4

Appendix B – Survey of attitudes toward statistics (SATS-36)

DIRECTIONS: The statements below are designed to identify your attitudes about statistics. Each item has 7 possible responses. The responses range from 1 (strongly disagree) through 4 (neutral) to 7 (strongly agree). If you have no opinion, choose response 4. Please read each statement. Mark the one response that most clearly represents your degree of agreement or disagreement with that statement. Try not to think too deeply about each response. Record your answer and move quickly to the next item. Please respond to all of the statements. CAUTION: You have only 20 MINUTES to finish this session.

	Strongly disagree <--> Neutral <--> Strongly agree						
1. I plan to complete all of my statistics assignments.	☐ 1	☐ 2	☐ 3	☐ 4	☐ 5	☐ 6	☐ 7
2. I plan to work hard in my statistics course.	☐ 1	☐ 2	☐ 3	☐ 4	☐ 5	☐ 6	☐ 7
3. I will like statistics.	☐ 1	☐ 2	☐ 3	☐ 4	☐ 5	☐ 6	☐ 7
4. I will feel insecure when I have to do statistics problems.	☐ 1	☐ 2	☐ 3	☐ 4	☐ 5	☐ 6	☐ 7
5. I will have trouble understanding statistics because of how I think.	☐ 1	☐ 2	☐ 3	☐ 4	☐ 5	☐ 6	☐ 7
6. Statistics formulas are easy to understand.	☐ 1	☐ 2	☐ 3	☐ 4	☐ 5	☐ 6	☐ 7
7. Statistics is worthless.	☐ 1	☐ 2	☐ 3	☐ 4	☐ 5	☐ 6	☐ 7
8. Statistics is a complicated subject.	☐ 1	☐ 2	☐ 3	☐ 4	☐ 5	☐ 6	☐ 7
9. Statistics should be a required part of my professional training.	☐ 1	☐ 2	☐ 3	☐ 4	☐ 5	☐ 6	☐ 7
10. Statistical skills will make me more employable.	☐ 1	☐ 2	☐ 3	☐ 4	☐ 5	☐ 6	☐ 7
11.I will have no idea of what's going on in this statistics course.	☐ 1	☐ 2	☐ 3	☐ 4	☐ 5	☐ 6	☐ 7

12. I am interested in being able to communicate statistical information to others.	☐ 1	☐ 2	☐ 3	☐ 4	☐ 5	☐ 6	☐ 7
13. Statistics is not useful to the typical professional.	☐ 1	☐ 2	☐ 3	☐ 4	☐ 5	☐ 6	☐ 7
14. I plan to study hard for every statistics test.	☐ 1	☐ 2	☐ 3	☐ 4	☐ 5	☐ 6	☐ 7
15. I will get frustrated going over statistics tests in class.	☐ 1	☐ 2	☐ 3	☐ 4	☐ 5	☐ 6	☐ 7
16.Statistical thinking is not applicable in my life outside my job.	☐ 1	☐ 2	☐ 3	☐ 4	☐ 5	☐ 6	☐ 7
17.I use statistics in my everyday life.	☐ 1	☐ 2	☐ 3	☐ 4	☐ 5	☐ 6	☐ 7
18. I will be under stress during statistics class.	☐ 1	☐ 2	☐ 3	☐ 4	☐ 5	☐ 6	☐ 7
19. I will enjoy taking statistics courses.	☐ 1	☐ 2	☐ 3	☐ 4	☐ 5	☐ 6	☐ 7
20. I am interested in using statistics.	☐ 1	☐ 2	☐ 3	☐ 4	☐ 5	☐ 6	☐ 7
21. Statistics conclusions are rarely presented in everyday life.	☐ 1	☐ 2	☐ 3	☐ 4	☐ 5	☐ 6	☐ 7
22. Statistics is a subject quickly learned by most people	☐ 1	☐ 2	☐ 3	☐ 4	☐ 5	☐ 6	☐ 7
23.I am interested in understanding statistical information.	☐ 1	☐ 2	☐ 3	☐ 4	☐ 5	☐ 6	☐ 7
24.Learning statistics requires a great deal of discipline.	☐ 1	☐ 2	☐ 3	☐ 4	☐ 5	☐ 6	☐ 7
25.I will have no application for statistics in my profession.	☐ 1	☐ 2	☐ 3	☐ 4	☐ 5	☐ 6	☐ 7
26.I will make a lot of math errors in statistics.	☐ 1	☐ 2	☐ 3	☐ 4	☐ 5	☐ 6	☐ 7
27.I plan to attend every statistics class session.	☐ 1	☐ 2	☐ 3	☐ 4	☐ 5	☐ 6	☐ 7
28.I am scared by statistics.	☐ 1	☐ 2	☐ 3	☐ 4	☐ 5	☐ 6	☐ 7

29.I am interested in learning statistics.	☐ 1	☐ 2	☐ 3	☐ 4	☐ 5	☐ 6	☐ 7
30.Statistics involves massive computations.	☐ 1	☐ 2	☐ 3	☐ 4	☐ 5	☐ 6	☐ 7
31.I can learn statistics.	☐ 1	☐ 2	☐ 3	☐ 4	☐ 5	☐ 6	☐ 7
32. I will understand statistics equations.	☐ 1	☐ 2	☐ 3	☐ 4	☐ 5	☐ 6	☐ 7
33. Statistics is irrelevant in my life.	☐ 1	☐ 2	☐ 3	☐ 4	☐ 5	☐ 6	☐ 7
34. Statistics is highly technical.	☐ 1	☐ 2	☐ 3	☐ 4	☐ 5	☐ 6	☐ 7
35. I will find it difficult to understand statistical concepts.	☐ 1	☐ 2	☐ 3	☐ 4	☐ 5	☐ 6	☐ 7
36. Most people have to learn a new way of thinking to do statistics.	☐ 1	☐ 2	☐ 3	☐ 4	☐ 5	☐ 6	☐ 7

Appendix C – Pre-test questions for the pilot study

Question 1. (field 1-3)

1) The status quo or prevailing viewing about the population is stated as the _____ hypothesis, which is denoted by _____.

A) Null; H0
B) Alternative; H1
C) Null; H1
D) Alternative; H0

2) How confident do you feel about the answer?
A) Very confident
B) Somewhat Confident
C) Not confident
D) Not confident at all

Question2. (field 4-6)

1) Give the statistical notation to denote the chance of making a Type-I error.
A) None of the above
B) *p*-value
C) α
D) β

2) How confident do you feel about the answer?
A) Very confident
B) Somewhat Confident
C) Not confident
D) Not confident at all

Question3. (field 7-9)

1) The significance level α is the chance that H_0 is true.
A) True
B) False

2) How confident do you feel about the answer?
A) Very confident
B) Somewhat Confident
C) Not confident

D) Not confident at all

Question 4. (field 10-12)

1) Which one of the following assumptions must be made in order to compute the *p*-value?

A) The null hypothesis is true
B) The alternative hypothesis is true
C) None of above

2) How confident do you feel about the answer?
A) Very confident
B) Somewhat Confident
C) Not confident
D) Not confident at all

Question5. (field 13-15)

1) Give the statistical terminology that represents the chance of accepting H0 when in fact H1 is true.

A) *p*-value
B) α
C) β
D) None of above

2) How confident do you feel about the answer?
A) Very confident
B) Somewhat Confident
C) Not confident
D) Not confident at all

Question6. (field 16-18)

1) The *p*-value can be determined without observing data.
A) True
B) False

2) How confident do you feel about the answer?
A) Very confident
B) Somewhat Confident
C) Not confident
D) Not confident at all

Question7. (field 19-21)

1) If the data are very likely to be observed under the assumption that the alternative hypothesis is true, it is appropriate for us to say the data is statistically significant.
A) True
B) False

2) How confident do you feel about the answer?
A) Very confident
B) Somewhat Confident
C) Not confident
D) Not confident at all

Question 8. (field 22-24)

1) Assume that in a certain testing of a study, the *p*-value of the observed data has been shown to be smaller than the level of significance (α). This implies that

A) The alternative hypothesis is being supported by the observed data
B) The null hypothesis is being supported by the observed data
C) None of above

2) How confident do you feel about the answer?
A) Very confident
B) Somewhat Confident
C) Not confident
D) Not confident at all

Question9. (field 25-27)

1) In hypothesis testing, which of the following statements are always true?

A) The *p*-value is greater than the significance level (α).
B) The *p*-value is computed from the significance level (α).
C) The *p*-value is the parameter in the null hypothesis.
D) The *p*-value is a test statistic.
E) The *p*-value is a probability.

2) How confident do you feel about the answer?
A) Very confident
B) Somewhat Confident
C) Not confident
D) Not confident at all

Question10. (field 28-30)

1) What type of error is possible if we decide to support the alternative hypothesis when in fact the null hypothesis is true?
A) No error possible
B) Type I error
C) Both Type I & II
D) Type II error

2) How confident do you feel about the answer?
A) Very confident
B) Somewhat Confident
C) Not confident
D) Not confident at all

Question11. (field 31-33)

1) There are two boxes: Box A and Box B.

Box A has 10 blue balls and 5 red balls.
Box B has 1 blue ball and 8 red balls.

A box is picked at random and a ball is drawn at random from that box and shown to you. Based on the color of the ball, you must decide if the ball was picked from Box A or Box B. Consider the following hypothesis:

Null: The ball was picked from Box A
Alternate: The ball was picked from Box B

Consider the following decision rule: Reject H0 if the ball is red.
Suppose that the ball drawn is blue. Compute α, the chance of making Type I error.

A) 1/9
B) 2/3
C) 1/3
D) None of above

2) How confident do you feel about the answer?
A) Very confident
B) Somewhat Confident
C) Not confident
D) Not confident at all

Question12. (field 34-36)

1) There are two identical looking wallets A and B and here are descriptions of the contents:

Wallet A: One $1 bills, Five $5 bills, Ten $10 bills
Wallet B: Eight $1 bills, Seven $5 bills, One $10 bills

A wallet is picked at random and exactly one bill is drawn at random from that wallet and shown to you. Based on the observation, you must decide which wallet the bill was drawn from.
Consider the following hypothesis.

H0: The bill is drawn from wallet A
H1: The bill is drawn from wallet B

Consider the decision rule: Reject H0 if the bill drawn is <=$5, otherwise accept H0.
The bill drawn is $1.
Compute β, the type II error.

A) 1/16
B) 1/2
C) 3/8
D) None of above

2) How confident do you feel about the answer?
A) Very confident
B) Somewhat Confident
C) Not confident
D) Not confident at all

Question 13. (field 37-39)

1) There are two identical looking wallets A and B and here are descriptions of the contents:

Wallet A: Eight $1 bills, Seven $5 bills, One $10 bills
Wallet B: One $1 bills, Five $5 bills, Ten $10 bills

A wallet is picked at random and exactly one bill is drawn at random from that wallet and shown to you. Based on the observation, you must decide which wallet the bill was drawn from.
Consider the following hypothesis.

H0: The bill is drawn from wallet A
H1: The bill is drawn from wallet B

Consider the decision rule: Reject H0 if the bill drawn is >=$5, otherwise accept H0.
For the given decision rule, the rejection region is

A) Two sided
B) One-sided to the right
C) One-sided to the left

2) How confident do you feel about the answer?
A) Very confident
B) Somewhat Confident
C) Not confident
D) Not confident at all

Question 14. (field 40-42)

1) There are two identical looking wallets A and B and here are descriptions of the contents:

Wallet A: One $1 bills, Five $5 bills, Ten $10 bills
Wallet B: Eight $1 bills, Seven $5 bills, One $10 bills

A wallet is picked at random and exactly one bill is drawn at random from that wallet and shown to you. Based on the observation, you must decide which wallet the bill was drawn from.
Consider the following hypothesis.

H0: The bill is drawn from wallet A
H1: The bill is drawn from wallet B

Consider the decision rule: Reject H0 if the bill drawn is <=$5, otherwise accept H0.
For the given decision rule, if the bill picked up is $5, compute the *p*-value.

A) 1/16
B) 3/8
C) 15/16
D) None of above

2) How confident do you feel about the answer?
A) Very confident
B) Somewhat Confident
C) Not confident

D) Not confident at all

Question 15 (field 43-45)

1) A box contains many coins and all of them look identical. However only some coins are fair - when you toss a fair coin, the probability of getting heads is 1/2. The other coins are biased and assume that when a biased coin is tossed, the probability of getting heads is 2/3. We are given one coin from the box and asked to determine if it is a fair coin, by tossing the coin exactly once and basing our decision on the outcome.

Consider the null hypothesis, H0: The tossed coin is a fair coin.
Consider the decision rule: Reject H0 if we get heads.

Calculate α, the chance of making Type I error.

A) cannot be determined from the given information.
B) 1/3
C) 1/2
D) 2/3

2) How confident do you feel about the answer?
A) Very confident
B) Somewhat Confident
C) Not confident
D) Not confident at all

Question 16.(field 46-48)

1) An educational study claims that girls perform better than boys in language arts. Consider the following hypotheses to be tested: H0: Girls' performance in language arts is the same as boys', versus H1: Girls' performance in language arts is better than boys'. If the study was statically significant at a significance of 0.05, which of the following value could be the *p*-value of this study?

A) 0.5
B) 1.0
C) 0.01
D) 0.1

2) How confident do you feel about the answer?
A) Very confident
B) Somewhat Confident
C) Not confident
D) Not confident at all

Appendix D – Post-test questions for the pilot study

Question 1. (field 1-3)

1) The status quo or prevailing viewing about the population is stated as the _____ hypothesis, which is denoted by _____.

A) Null; H0
B) Alternative; H1
C) Null; H1
D) Alternative; H0

2) How confident do you feel about the answer?
A) Very confident
B) Somewhat Confident
C) Not confident
D) Not confident at all

Question 2. (field 4-6)

1) Brian wants to examine the relationship between gender and GPA. Let Pf denotes the GPA of female students in the population. Suppose the GPA of male students in the population is known to be 3.20. We wish to assess if the population GPA of female students is greater than the population GPA of male students. State the hypotheses to be tested in terms of Pf.

A) H0: Pf >3.2 versus H1: Pf<3.2
B) H0: Pf =3.2 versus H1: Pf>3.2
C) H0: Pf <3.2 versus H1: Pf>3.2
D) H0: Pf >3.2 versus H1: Pf=3.2

2) How confident do you feel about the answer?
A) Very confident
B) Somewhat Confident
C) Not confident
D) Not confident at all

Question 3. (field 7-9)

1) Which one of the following assumptions must be made in order to compute the *p*-value?

A) The null hypothesis is true
B) The alternative hypothesis is true
C) None of above

2) How confident do you feel about the answer?
A) Very confident
B) Somewhat Confident
C) Not confident
D) Not confident at all

Question 4. (field 10-12)

1) Assume that in a certain testing of a study, the *p*-value of the observed data has been shown to be smaller than the level of significance (α). This implies that

A) The alternative hypothesis is being supported by the observed data
B) The null hypothesis is being supported by the observed data
C) None of above

2) How confident do you feel about the answer?
A) Very confident
B) Somewhat Confident
C) Not confident
D) Not confident at all

Question5. (field 13-15)

1) Give the statistical terminology that represents the chance of accepting H0 when in fact H1 is true.

A) *p*-value
B) α
C) β
D) None of above

2) How confident do you feel about the answer?
A) Very confident
B) Somewhat Confident
C) Not confident
D) Not confident at all

Question6. (field 16-18)

1) In hypothesis testing, which of the following statements are always true?

A) The *p*-value is greater than the significance level (α).
B) The *p*-value is computed from the significance level (α).
C) The *p*-value is the parameter in the null hypothesis.
D) The *p*-value is a test statistic.
E) The *p*-value is a probability.

2) How confident do you feel about the answer?
A) Very confident
B) Somewhat Confident
C) Not confident
D) Not confident at all

Question7. (field 19-21)

1) For a fixed sample size, the chance of a Type I error (α) can be decreased by increasing the Type II error (β).
A) True
B) False

2) How confident do you feel about the answer?
A) Very confident
B) Somewhat Confident
C) Not confident
D) Not confident at all

Question8. (field 22-24)

1) If the data are very likely to be observed under the assumption that the alternative hypothesis is true, it is appropriate for us to say the data is statistically significant.
A) True
B) False

2) How confident do you feel about the answer?
A) Very confident
B) Somewhat Confident
C) Not confident
D) Not confident at all

Question9. (field 25-27)

1) There are two boxes: Box A and Box B.

Box A has 10 blue balls and 5 red balls.
Box B has 1 blue ball and 8 red balls.

A box is picked at random and a ball is drawn at random from that box and shown to you. Based on the color of the ball, you must decide if the ball was picked from Box A or Box B. Consider the following hypothesis:

Null: The ball was picked from Box A
Alternate: The ball was picked from Box B

Consider the following decision rule: Reject H0 if the ball is red.
Suppose that the ball drawn is blue. Compute α, the chance of making Type I error.

A) 1/9
B) 2/3
C) 1/3
D) None of above

2) How confident do you feel about the answer?
A) Very confident
B) Somewhat Confident
C) Not confident
D) Not confident at all

Question10. (field 28-30)

1) There are two identical looking wallets A and B and here are descriptions of the contents:

Wallet A: One $1 bills, Five $5 bills, Ten $10 bills
Wallet B: Eight $1 bills, Seven $5 bills, One $10 bills

A wallet is picked at random and exactly one bill is drawn at random from that wallet and shown to you. Based on the observation, you must decide which wallet the bill was drawn from.
Consider the following hypothesis.

H0: The bill is drawn from wallet A
H1: The bill is drawn from wallet B

Consider the decision rule: Reject H0 if the bill drawn is <=$5, otherwise accept H0.
The bill drawn is $1.

Compute α, the chance of making Type I error.

A) 1/16
B) 1/2
C) 3/8
D) None of above

2) How confident do you feel about the answer?
A) Very confident
B) Somewhat Confident
C) Not confident
D) Not confident at all

Question11. (field 31-33)

1) There are two boxes: Box A and Box B.

Box A has 10 blue balls and 5 red balls.
Box B has 1 blue ball and 8 red balls.

A box is picked at random and a ball is drawn at random from that box and shown to you. Based on the color of the ball, you must decide if the ball was picked from Box A or Box B. Consider the following hypothesis:

Null: The ball was picked from Box A
Alternate: The ball was picked from Box B

Consider the following decision rule: Reject H0 if the ball is red.
Suppose that the ball drawn is blue. Compute β, the type II error.

A) 1/3
B) 8/9
C) 1/9
D) None of above

2) How confident do you feel about the answer?
A) Very confident
B) Somewhat Confident
C) Not confident
D) Not confident at all

Question 12. (field 34-36)

1) There are two identical looking wallets A and B and here are descriptions of the contents:

Wallet A: Eight $1 bills, Seven $5 bills, One $10 bills
Wallet B: One $1 bills, Five $5 bills, Ten $10 bills

A wallet is picked at random and exactly one bill is drawn at random from that wallet and shown to you. Based on the observation, you must decide which wallet the bill was drawn from.
Consider the following hypothesis.

H0: The bill is drawn from wallet A
H1: The bill is drawn from wallet B

Consider the decision rule: Reject H0 if the bill drawn is >=$5, otherwise accept H0.
For the given decision rule, the rejection region is

A) Two sided
B) One-sided to the right
C) One-sided to the left

2) How confident do you feel about the answer?
A) Very confident
B) Somewhat Confident
C) Not confident
D) Not confident at all

Question 13. (field 37-39)

1) There are two identical looking wallets A and B and here are descriptions of the contents:

Wallet A: One $1 bills, Five $5 bills, Ten $10 bills
Wallet B: Eight $1 bills, Seven $5 bills, One $10 bills

A wallet is picked at random and exactly one bill is drawn at random from that wallet and shown to you. Based on the observation, you must decide which wallet the bill was drawn from.
Consider the following hypothesis.

H0: The bill is drawn from wallet A
H1: The bill is drawn from wallet B

Consider the decision rule: Reject H0 if the bill drawn is <=$5, otherwise accept H0.
For the given decision rule, if the bill picked up is $5, compute the *p*-value.

A) 1/16
B) 3/8
C) 15/16
D) None of above

2) How confident do you feel about the answer?
A) Very confident
B) Somewhat Confident
C) Not confident
D) Not confident at all

Question 14.(field 40-42)

1) A box contains many coins and all of them look identical. However only some coins are fair - when you toss a fair coin, the probability of getting heads is 1/2. The other coins are biased and assume that when a biased coin is tossed, the probability of getting heads is 2/3. We are given one coin from the box and asked to determine if it is a fair coin, by tossing the coin exactly once and basing our decision on the outcome.

Consider the null hypothesis, H0: The tossed coin is a fair coin.
Consider the decision rule: Reject H0 if we get heads.

Calculate α, the chance of making Type I error.

A) cannot be determined from the given information.
B) 1/3
C) 1/2
D) 2/3

2) How confident do you feel about the answer?
A) Very confident
B) Somewhat Confident
C) Not confident
D) Not confident at all

Question 15.(field 43-45)

1) A box contains many coins and all of them look identical. However only some coins are fair - when you toss a fair coin, the probability of getting heads is 1/2. The other coins are biased and assume that when a biased coin is tossed, the probability of getting heads

is 2/3. We are given one coin from the box and asked to determine if it is a fair coin, by tossing the coin exactly once and basing our decision on the outcome.

Consider the null hypothesis, H0: The tossed coin is a fair coin.
Consider the decision rule: Reject H0 if we get heads.

Calculate β, the chance of making Type II error.

A) cannot be determined from the given information.
B) 1/3
C) 1/2
D) 2/3

2) How confident do you feel about the answer?
A) Very confident
B) Somewhat Confident
C) Not confident
D) Not confident at all

Question 16.(field 46-48)

1) An educational study claims that girls perform better than boys in language arts. Consider the following hypotheses to be tested: H0: Girls' performance in language arts is the same as boys', versus H1: Girls' performance in language arts is better than boys'. If the study was NOT statically significant at a significance of 0.05 (α=0.05), which of the following value could be the *p*-value of this study?

A) 0.1
B) 0.01
C) 0.001
D) 0.0001

2) How confident do you feel about the answer?
A) Very confident
B) Somewhat Confident
C) Not confident
D) Not confident at all

Appendix E – Program perception survey for the pilot study

Please answer the following questions based on your perception of the program.

	Strongly disagree <-> Neutral <-> Strongly agree
1. The use of the multimedia program helps my understanding of statistics.	☐ 1 ☐ 2 ☐ 3 ☐ 4 ☐ 5 ☐ 6 ☐ 7
2. Comparing the use of textbook, the use of the multimedia program works better in developing my understanding of statistics.	☐ 1 ☐ 2 ☐ 3 ☐ 4 ☐ 5 ☐ 6 ☐ 7
3. It's easier for me to get the statistical ideas from the multimedia program than the textbook.	☐ 1 ☐ 2 ☐ 3 ☐ 4 ☐ 5 ☐ 6 ☐ 7

Appendix F – Pre-test questions for the main study

Question 1. (field 1-8)

There are two boxes: Box A and Box B.

Box A has <u>10</u> blue balls and <u>5</u> red balls.
Box B has <u>1</u> blue ball and <u>8</u> red balls.

A box is picked at random and a ball is drawn at random from that box and shown to you. Based on the color of the ball, you must decide if the ball was picked from Box A or Box B. Consider the following hypothesis:

Null (H0): The ball was picked from <u>Box A</u>
Alternate (H1): The ball was picked from <u>Box B</u>

Consider the following decision rule: <u>Reject H0 if the ball is red</u>.
The ball drawn is <u>red</u>.

Please compute <u>α, the type I error</u>, and answer the following questions.

1.1.1 The above statement mentioned α (the type I error). Which one of the following is correct based on your understanding to α (the type I error)?
A) Give the statistical terminology that represents the chance of accepting H0 when in fact H1 is true.
B) Give the statistical terminology that represents the chance of rejecting H0 when in fact H0 is true.
C) None of above

1.1.2 How confident do you feel about your answer for 1.1.1?
A) No confidence
B) A little confidence
C) Some confidence
D) Strong confidence
E) Complete confidence

1.2.1 Which of the following distribution should be used to compute α (the type I error) in this case?
A) The Box A distribution
B) The Box B distribution
C) None of above

1.2.2 How confident do you feel about your answer for 1.2.1?
A) No confidence
B) A little confidence
C) Some confidence
D) Strong confidence
E) Complete confidence

1.3.1 Select a correct value for α, the type I error.
A) 1/9
B) 2/3
C) 1/3
D) None of above

1.3.2 How confident do you feel about your answer for 1.3.1?
A) No confidence
B) A little confidence
C) Some confidence
D) Strong confidence
E) Complete confidence

1.4.1 Please select a correct interpretation for α, the type I error.
A) α=1/5 means that if H0 is rejected 5 times then will be 1 error.
B) α=1/5 means that if H0 is accepted 5 times then will be 1 error.
C) None of above

1.4.2 How confident do you feel about your answer for 1.4.1?
A) No confidence
B) A little confidence
C) Some confidence
D) Strong confidence
E) Complete confidence

Question2. (field 9-16)

There are two identical looking wallets A and B and here are descriptions of the contents:

Wallet A: <u>Two</u> $1 bills, <u>Five</u> $5 bills, <u>Ten</u> $10 bills
Wallet B: <u>Eight</u> $1 bills, <u>Seven</u> $5 bills, <u>One</u> $10 bills

A wallet is picked at random and exactly one bill is drawn at random from that wallet and shown to you. Based on the observation, you must decide which wallet the bill was drawn from.
Consider the following hypothesis.

H0: The bill is drawn from <u>Wallet A</u>
H1: The bill is drawn from <u>Wallet B</u>

Consider the decision rule: <u>Reject H0 if the bill drawn is <=$5</u>, otherwise accept H0. The bill drawn is <u>$1</u>.

Please compute <u>β, the type II error</u>, and answer the following questions.

2.1.1 The above statement mentioned β (the type II error). Which one of the following is correct based on your understanding to β (the type II error)?
A) Give the statistical terminology that represents the chance of accepting H0 when in fact H1 is true.
B) Give the statistical terminology that represents the chance of rejecting H0 when in fact H0 is true.
C) None of above

2.1.2 How confident do you feel about your answer for 2.1.1?
A) No confidence
B) A little confidence
C) Some confidence
D) Strong confidence
E) Complete confidence

2.2.1 Which of the following distribution should be used to compute β (the type II error) in this case?
A) The Wallet A distribution
B) The Wallet B distribution
C) None of above

2.2.2 How confident do you feel about your answer for 2.2.1?
A) No confidence
B) A little confidence
C) Some confidence
D) Strong confidence
E) Complete confidence

2.3.1 Select a correct value for β, the type II error.

A) 1/8

B) 1/16
C) 3/8
D) None of above

2.3.2 How confident do you feel about your answer for 2.3.1?
A) No confidence
B) A little confidence
C) Some confidence
D) Strong confidence
E) Complete confidence

2.4.1 Please select a correct interpretation for β, the type II error.
A) β=1/5 means that if H0 is rejected 5 times then will be 1 error.
B) β=1/5 means that if H0 is accepted 5 times then will be 1 error.
C) None of above

2.4.2 How confident do you feel about your answer for 2.4.1?
A) No confidence
B) A little confidence
C) Some confidence
D) Strong confidence
E) Complete confidence

Question3. (field 17-24)

There are two identical looking wallets A and B and here are descriptions of the contents:

Wallet A: <u>Eight</u> $1 bills, <u>Seven</u> $5 bills, <u>One</u> $10 bills
Wallet B: <u>One</u> $1 bills, <u>Five</u> $5 bills, <u>Ten</u> $10 bills

A wallet is picked at random and exactly one bill is drawn at random from that wallet and shown to you. Based on the observation, you must decide which wallet the bill was drawn from.

Consider the following hypothesis.

H0: The bill is drawn from <u>Wallet A</u>
H1: The bill is drawn from <u>Wallet B</u>

Consider the decision rule: <u>Reject H0 if the bill drawn is >=$5</u>, otherwise accept H0. The bill drawn is <u>$5</u>.

Please compute the *p*-value and answer the following questions.

3.1.1 The above statement mentioned *p*-value. Which one of the following is correct based on your understanding to *p*-value?
A) The *p*-value is a probability of how much evidence we have against the alternative hypothesis
B) The *p*-value needs to be determined with observing data.
C) The *p*-value is a chance, computed under the assumption that H1 is true, of getting the observed value plus the chance of getting all of the more extreme values
D) None of above

3.1.2 How confident do you feel about your answer for 3.1.1?
A) No confidence
B) A little confidence
C) Some confidence
D) Strong confidence
E) Complete confidence

3.2.1 Which of the following distribution should be used to compute the *p*-value in this case?
A) The Wallet A distribution
B) The Wallet B distribution
C) None of above

3.2.2 How confident do you feel about your answer for 3.2.1?
A) No confidence
B) A little confidence
C) Some confidence
D) Strong confidence
E) Complete confidence

2.3.1 Select a correct value for the *p*-value.

A) 1/16
B) 1/2
C) 15/16
D) None of above

3.3.2 How confident do you feel about your answer for 3.3.1?
A) No confidence
B) A little confidence
C) Some confidence
D) Strong confidence
E) Complete confidence

3.4.1 Please select a correct interpretation for the *p*-value in the case. Given the *p*-value is 1/5.
A) The chance getting a voucher value of $5 or more from bag A is 1/5.
B) The chance getting a voucher value of $5 or less from bag A is 1/5.
C) The chance getting a voucher value of $5 or more from bag B is 1/5.
D) The chance getting a voucher value of $5 or less from bag B is 1/5.

3.4.2 How confident do you feel about your answer for 3.4.1?
A) No confidence
B) A little confidence
C) Some confidence
D) Strong confidence
E) Complete confidence

Question4. (field 15-30)

A researcher claims that girls perform better than boys in language arts. Consider the following hypotheses tested: <u>H0</u>: Girls' performance in language arts is the <u>same</u> as boys', versus <u>H1</u>: Girls' performance in language arts is <u>better</u> than boys'. If the study was statistically significant at a <u>significance of 0.05 (*p*-value < α ; α =0.05)</u>, please answer the following questions.

4.1.1 The statistical notation α is for
A) The chance of making a Type-I error
B) The chance of making a Type-II error
C) Null hypothesis
D) Alternative hypothesis

4.1.2 How confident do you feel about your answer for 4.1.1?
A) No confidence
B) A little confidence
C) Some confidence
D) Strong confidence
E) Complete confidence

4.2.1 Give the statistical terminology that represents the chance of accepting H0 when in fact it is false.
A) p - value
B) β
C) Level of significance denoted by α
D) None of the above

4.2.2 How confident do you feel about your answer for 4.2.1?
A) No confidence
B) A little confidence
C) Some confidence
D) Strong confidence
E) Complete confidence

4.3.1 If the observed data are stated to be statistically significant, then which of the following hypothesis is being supported?
A) Both the hypotheses
B) Neither of the hypotheses
C) Alternate Hypothesis
D) Null Hypothesis

4.3.2 How confident do you feel about your answer for 4.3.1?
A) No confidence
B) A little confidence
C) Some confidence
D) Strong confidence
E) Complete confidence

4.4.1 If the *p*-value of the observed data has been shown to be smaller than the level of significance α. This implies that
A) the null hypothesis is being supported by the observed data.
B) neither of the hypothesis is being supported.
C) the alternate hypothesis is being supported by the observed data
D) None of the above

4.4.2 How confident do you feel about your answer for 4.4.1?
A) No confidence
B) A little confidence
C) Some confidence
D) Strong confidence
E) Complete confidence

4.5.1. If the study was statically significant at a significance of 0.05, which of the following value could be the *p*-value of this study?
A) 0.001
B) 1.0
C) 0.01
D) 0.1

4.5.2 How confident do you feel about your answer for 4.5.1?
A) No confidence

B) A little confidence
C) Some confidence
D) Strong confidence
E) Complete confidence

4.6.1. If the study was statically significant at a significance of 0.01, which of the following value could be the *p*-value of this study?

A) 0.5
B) 0.05
C) 0.005
D) 0.1

4.6.2 How confident do you feel about your answer for 4.6.1?
A) No confidence
B) A little confidence
C) Some confidence
D) Strong confidence
E) Complete confidence

4.7.1 Which of the following sentences is consistent with the above case?
I. The observed data, or data even more extreme, are unlikely to occur if the null hypothesis is true.
II. Alternate hypothesis is being supported by the observed data.
III. Type II error is possible
A) II only
B) I only
C) I, II & III
D) I & II only

4.7.2 How confident do you feel about your answer for 4.7.1?
A) No confidence
B) A little confidence
C) Some confidence
D) Strong confidence
E) Complete confidence

4.8.1 Is it possible for this study to commit an error in this case? Please describe the error that the study could make and show the possibility.
A) Yes. Yes. It is possible for the researcher to make errors. The type I error (α) 0.05 means there will be 5 errors over 100 tests.
B) Yes. It is possible for the researcher to make errors. The type I error (α) in this case is 0.05, which means that if H1 is rejected 100 times then will be 5 errors.

C) No. It is not possible for the researcher to make any mistake since the value of α (0.05) is very small.
D) No. It is not possible for the researcher to make any mistake since the testing result is significant.

4.8.2 How confident do you feel about your answer for 4.8.1?
A) No confidence
B) A little confidence
C) Some confidence
D) Strong confidence
E) Complete confidence

Appendix G– Post-test questions for the main study

Question 1. (field 1-8)

There are two boxes: Box A and Box B.

Box A has <u>10</u> blue balls and <u>2</u> red balls.
Box B has <u>1</u> blue ball and <u>11</u> red balls.

A box is picked at random and a ball is drawn at random from that box and shown to you. Based on the color of the ball, you must decide if the ball was picked from Box A or Box B. Consider the following hypothesis:

Null (H0): The ball was picked from <u>Box A</u>
Alternate (H1): The ball was picked from <u>Box B</u>

Consider the following decision rule: Reject H0 if the ball is red.
The ball drawn is <u>red</u>.

Please compute <u>α, the type I error</u>, and answer the following questions.

1.1.1 The above statement mentioned α (the type I error). Which one of the following is correct based on your understanding to α (the type I error)?
A) Give the statistical terminology that represents the chance of accepting H0 when in fact H1 is true.
B) Give the statistical terminology that represents the chance of rejecting H0 when in fact H0 is true.
C) None of above

1.1.2 How confident do you feel about your answer for 1.1.1?
A) No confidence
B) A little confidence
C) Some confidence
D) Strong confidence
E) Complete confidence

1.2.1 Which of the following distribution should be used to compute α (the type I error) in this case?
A) The Box A distribution
B) The Box B distribution
C) None of above

1.2.2 How confident do you feel about your answer for 1.2.1?
A) No confidence
B) A little confidence
C) Some confidence
D) Strong confidence
E) Complete confidence

1.3.1 Select a correct value for α, the type I error.
A) 1/9
B) 2/3
C) 1/3
D) None of above

1.3.2 How confident do you feel about your answer for 1.3.1?
A) No confidence
B) A little confidence
C) Some confidence
D) Strong confidence
E) Complete confidence

1.4.1 Please select a correct interpretation for α, the type I error.
A) α=1/5 means that if H0 is rejected 5 times then will be 1 error.
B) α=1/5 means that if H0 is accepted 5 times then will be 1 error.
C) None of above

1.4.2 How confident do you feel about your answer for 1.4.1?
A) No confidence
B) A little confidence
C) Some confidence
D) Strong confidence
E) Complete confidence

Question2. (field 9-16)

There are two identical looking wallets A and B and here are descriptions of the contents:

Wallet A: <u>One</u> $1 bills, <u>Five</u> $5 bills, <u>Ten</u> $10 bills
Wallet B: <u>Eight</u> $1 bills, <u>Six</u> $5 bills, <u>Two</u> $10 bills

A wallet is picked at random and exactly one bill is drawn at random from that wallet and shown to you. Based on the observation, you must decide which wallet the bill was drawn from.
Consider the following hypothesis.

H0: The bill is drawn from <u>Wallet A</u>
H1: The bill is drawn from <u>Wallet B</u>

Consider the decision rule: <u>Reject H0 if the bill drawn is <=$5</u>, otherwise accept H0.
The bill drawn is <u>$1</u>.

Please compute <u>β, the type II error</u>, and answer the following questions.

2.1.1 The above statement mentioned β (the type II error). Which one of the following is correct based on your understanding to β (the type II error)?
A) Give the statistical terminology that represents the chance of accepting H0 when in fact H1 is true.
B) Give the statistical terminology that represents the chance of rejecting H0 when in fact H0 is true.
C) None of above

2.1.2 How confident do you feel about your answer for 2.1.1?
A) No confidence
B) A little confidence
C) Some confidence
D) Strong confidence
E) Complete confidence

2.2.1 Which of the following distribution should be used to compute β (the type II error) in this case?
A) The Wallet A distribution
B) The Wallet B distribution
C) None of above

2.2.2 How confident do you feel about your answer for 2.2.1?
A) No confidence
B) A little confidence
C) Some confidence
D) Strong confidence
E) Complete confidence

2.3.1 Select a correct value for β, the type II error.

A) 1/8

B) 1/16
C) 3/8
D) None of above

2.3.2 How confident do you feel about your answer for 2.3.1?
A) No confidence
B) A little confidence
C) Some confidence
D) Strong confidence
E) Complete confidence

2.4.1 Please select a correct interpretation for β, the type II error.
A) β=1/5 means that if H0 is rejected 5 times then will be 1 error.
B) β=1/5 means that if H0 is accepted 5 times then will be 1 error.
C) None of above

2.4.2 How confident do you feel about your answer for 2.4.1?
A) No confidence
B) A little confidence
C) Some confidence
D) Strong confidence
E) Complete confidence

Question3. (field 17-24)

There are two identical looking wallets A and B and here are descriptions of the contents:

Wallet A: <u>Eight</u> $1 bills, <u>Six</u> $5 bills, <u>Two</u> $10 bills
Wallet B: <u>One</u> $1 bills, <u>Five</u> $5 bills, <u>Ten</u> $10 bills

A wallet is picked at random and exactly one bill is drawn at random from that wallet and shown to you. Based on the observation, you must decide which wallet the bill was drawn from.
Consider the following hypothesis.

H0: The bill is drawn from <u>Wallet A</u>
H1: The bill is drawn from <u>Wallet B</u>

Consider the decision rule: <u>Reject H0 if the bill drawn is >=$5</u>, otherwise accept H0.
The bill drawn is <u>$5</u>.

Please compute the <u>*p*-value</u> and answer the following questions.

3.1.1 The above statement mentioned *p*-value. Which one of the following is correct based on your understanding to *p*-value?
A) The *p*-value is a probability of how much evidence we have against the alternative hypothesis
B) The *p*-value needs to be determined with observing data.
C) The *p*-value is a chance, computed under the assumption that H1 is true, of getting the observed value plus the chance of getting all of the more extreme values
D) None of above

3.1.2 How confident do you feel about your answer for 3.1.1?
A) No confidence
B) A little confidence
C) Some confidence
D) Strong confidence
E) Complete confidence

3.2.1 Which of the following distribution should be used to compute the *p*-value in this case?
A) The bag A distribution
B) The bag B distribution
C) None of above

3.2.2 How confident do you feel about your answer for 3.2.1?
A) No confidence
B) A little confidence
C) Some confidence
D) Strong confidence
E) Complete confidence

2.3.1 Select a correct value for the *p*-value.

A) 1/16
B) 1/2
C) 15/16
D) None of above

3.3.2 How confident do you feel about your answer for 3.3.1?
A) No confidence
B) A little confidence
C) Some confidence
D) Strong confidence
E) Complete confidence

3.4.1 Please select a correct interpretation for the *p*-value in the case. Given the *p*-value is 1/5.
A) The chance getting a voucher value of $5 or more from bag A is 1/5.
B) The chance getting a voucher value of $5 or less from bag A is 1/5.
C) The chance getting a voucher value of $5 or more from bag B is 1/5.
D) The chance getting a voucher value of $5 or less from bag B is 1/5.

3.4.2 How confident do you feel about your answer for 3.4.1?
A) No confidence
B) A little confidence
C) Some confidence
D) Strong confidence
E) Complete confidence

Question4. (field 15-30)

A researcher claims that girls perform better than boys in language arts. Consider the following hypotheses tested: <u>H0</u>: Girls' performance in language arts is the <u>same</u> as boys', versus <u>H1</u>: Girls' performance in language arts is <u>better</u> than boys'. If the study was statistically significant at a <u>significance of 0.05 (*p*-value < α ; α =0.05)</u>, please answer the following questions.

4.1.1 The statistical notation α is for
A) The chance of making a Type-I error
B) The chance of making a Type-II error
C) Null hypothesis
D) Alternative hypothesis

4.1.2 How confident do you feel about your answer for 4.1.1?
A) No confidence
B) A little confidence
C) Some confidence
D) Strong confidence
E) Complete confidence

4.2.1 Give the statistical terminology that represents the chance of accepting H0 when in fact it is false.
A) p - value
B) β
C) Level of significance denoted by α
D) None of the above
4.2.2 How confident do you feel about your answer for 4.2.1?
A) No confidence
B) A little confidence

C) Some confidence
D) Strong confidence
E) Complete confidence

4.3.1 If the observed data are stated to be statistically significant, then which of the following hypothesis is being supported?
A) Both the hypotheses
B) Neither of the hypotheses
C) Alternate Hypothesis
D) Null Hypothesis

4.3.2 How confident do you feel about your answer for 4.3.1?
A) No confidence
B) A little confidence
C) Some confidence
D) Strong confidence
E) Complete confidence

4.4.1 If the *p*-value of the observed data has been shown to be smaller than the level of significance α. This implies that
A) the null hypothesis is being supported by the observed data.
B) neither of the hypothesis is being supported.
C) the alternate hypothesis is being supported by the observed data
D) None of the above

4.4.2 How confident do you feel about your answer for 4.4.1?
A) No confidence
B) A little confidence
C) Some confidence
D) Strong confidence
E) Complete confidence

4.5.1. If the study was statically significant at a significance of 0.05, which of the following value could be the *p*-value of this study?
A) 0.001
B) 1.0
C) 0.01
D) 0.1

4.5.2 How confident do you feel about your answer for 4.5.1?
A) No confidence
B) A little confidence
C) Some confidence
D) Strong confidence

E) Complete confidence

4.6.1. If the study was statically significant at a significance of 0.01, which of the following value could be the *p*-value of this study?

A) 0.5
B) 0.05
C) 0.005
D) 0.1

4.6.2 How confident do you feel about your answer for 4.6.1?
A) No confidence
B) A little confidence
C) Some confidence
D) Strong confidence
E) Complete confidence

4.7.1 Which of the following sentences is consistent with the above case?
I. The observed data, or data even more extreme, are unlikely to occur if the null hypothesis is true.
II. Alternate hypothesis is being supported by the observed data.
III. Type II error is possible
A) II only
B) I only
C) I, II & III
D) I & II only

4.7.2 How confident do you feel about your answer for 4.7.1?
A) No confidence
B) A little confidence
C) Some confidence
D) Strong confidence
E) Complete confidence

4.8.1 Is it possible for this study to commit an error in this case? Please describe the error that the study could make and show the possibility.
A) Yes. Yes. It is possible for the researcher to make errors. The type I error (α) 0.05 means there will be 5 errors over 100 tests.
B) Yes. It is possible for the researcher to make errors. The type I error (α) in this case is 0.05, which means that if H1 is rejected 100 times then will be 5 errors.
C) No. It is not possible for the researcher to make any mistake since the value of α (0.05) is very small.
D) No. It is not possible for the researcher to make any mistake since the testing result is significant.

4.8.2 How confident do you feel about your answer for 4.8.1?
A) No confidence
B) A little confidence
C) Some confidence
D) Strong confidence
E) Complete confidence

Appendix H – Program perception survey for the main study

1. I think this multimedia program can help me learn.
   Not at all Less Somewhat More All the time

2. I pay attention more if a teacher uses this multimedia program in his instruction.
   Not at all Less Somewhat More All the time

3. I better understand what is being taught if my teacher uses the multimedia program in his instruction.
   Not at all Less Somewhat More All the time

4. I remember more information if my teacher uses the multimedia program in his instruction.
   Not at all Less Somewhat More All the time

(Revised from Spaulding's Survey for Student Learning and Assessments in 2007)

Appendix I – Informed consent form for participants

IRB APPROVED ON: 08/21/2009 ***EXPIRES ON: 01/13/2010***
IRB # 2008-11-0022

Informed Consent Form for participants in phase 1

You are being asked to participate in a research study to receive credits for the class which you registered in department of educational psychology based on the **Subject Pool Requirement**. This form provides you with information about the study. The person in charge of this research will also describe this study to you and answer all of your questions. Please read the information below and ask any questions you might have before deciding whether or not to take part. Your participation is entirely voluntary. You can refuse or stop to participate to the study. However, based on the Subject Pool Requirement, an alternative assignment will be offered by those in charge of the subject pool for students not willing to participate. This assignment, given at semester's end, will be a written review of related research. The time needed to complete the assignment will be approximately the same as that for the research study participation. If you have additional questions about the Subject Pool requirement or have trouble registering through this website: http://edpsych.edb.utexas.edu/curriculum/SubjectPool/students/.

I. Title: Using the animated visual aid to facilitate the learning of introductory statistics

II. Investigators

Pei-Yu Wang, principal investigator	Instructional Technology	512-638-3759 sweetpeiyu@gmail.com
Brandon Vaughn, principal investigator	Educational Psychology	471-0792 brandon.vaughn@mail.utexas.edu
Min Liu, principal investigator	Instructional Technology	512-471-5211 mliu@mail.utexas.edu

III. Hypothesis, Research Questions, or Goals of the Project

The use of visual aids has the potential to enhance students' learning in statistics. However, not all visual aids perform equally well in students' learning process. This research project aims to explore the strength and limitation of visual aids. The goals of this project are:

1. to understand if students' individual characteristics impact their use of visual aids while learning statistics.
2. to understand how the visual aid impacts students while learning statistics.

IV. What are we asking of you?

We are asking you to help us determine the effectiveness of active learning and technology in learning statistics. You will be covering one module of learning, and taking a pre- and post-test for each module. It will take approximately 2 hours to complete the tasks. You are being asked to participate in one of the four following groups of this study:

1. Some of you will be asked to **learn statistics with the traditional Web-based instruction**.

2. Some of you will be asked to **learn statistics with an animation which you can control the sequences**.
3. Some of you will be asked to **learn statistics with an animation which you can control inputs and variables**.
4. Some of you will be asked to **learn statistics with an animation which you can practice and get feedback**.

If you are willing to participate, simply complete the tasks we ask of you. If you do not wish to participate, you are free to just not participate with no fear of adverse consequences. **This sheet is provided for your information only.**

V. Risks and Benefits

To the best of our knowledge, there are minimal risks involved in your participation in the project. We will not be retaining any personally identifiable information about you or your participation. We cannot foresee any personal benefits to you at this point other than the satisfaction of having helped us investigate exciting new ways of teaching statistics.

For more information: Please feel free to contact either Pei-Yu Wang (638-3759), Dr. Brandon Vaughn (471-0792) or Dr. Jody Jensen, at 232-2685, the chair of the IRB, to learn more about your rights in this matter.

VI. Confidentiality and Privacy Protections

The data resulting from your participation may be made available to other researchers in the future for research purposes not detailed within this consent form. In these cases, the data will contain no identifying information that could associate you with it, or with your participation in any study.
The records of this study will be stored securely and kept confidential. Authorized persons from The University of Texas at Austin, members of the Institutional Review Board, and (study sponsors, if any) have the legal right to review your research records and will protect the confidentiality of those records to the extent permitted by law. All publications will exclude any information that will make it possible to identify you as a subject. Throughout the study, the researchers will notify you of new information that may become available and that might affect your decision to remain in the study.
For more information: Please feel free to contact either Pei-Yu Wang (638-3759), Dr. Brandon Vaughn (471-0792) or Dr. Jody Jensen, at 232-2685, the chair of the IRB, to learn more about your rights in this matter.
If you are willing to participate, please sign the attached consent form and return it the researcher visiting your class.

Please retain this portion of the handout for your information.

Statement of Consent for <u>using the animated visual aid to facilitate the learning of introductory statistics</u>:

I have read the above information and have sufficient information to make a decision about participating in this study. I consent to participate in the study.

_____ Date: _____
Your Signature

_____ Date: _____
Signature of Person Obtaining Consent

_____ Date: _____
Signature of Investigator

Appendix J – Bonferroni correction for post hoc comparisons

Dependent Variable	Post hoc	(I) Group	(J) Group	Mean Difference (I-J)	Sig.
Achievement improvement on understanding	Tukey HSD	Static	Simple Animation	-1.27	.015(*)
			Input	-1.09	.045(*)
			Practice	-1.28	.012(*)
		Simple Animation	Static	1.27	.015(*)
			Input	.17	.976
			Practice	-.01	1.000
		Input	Static	1.09	.045(*)
			Simple Animation	-.17	.976
			Practice	-.18	.970
		Practice	Static	1.28	.012(*)
			Simple Animation	.01	1.000
			Input	.18	.970
	Bonferroni	Static	Simple Animation	-1.27	.017(*)
			Input	-1.09	.055
			Practice	-1.28	.014(*)
		Simple Animation	Static	1.27	.017(*)
			Input	.17	1.000
			Practice	-.01	1.000
		Input	Static	1.09	.055
			Simple Animation	-.17	1.000
			Practice	-.18	1.000
		Practice	Static	1.28	.014(*)
			Simple Animation	.01	1.000
			Input	.18	1.000

Based on observed means.
* The mean difference is significant at the .05 level.

Appendix K – Table for the effect size

	Treatment group	Static	Simple animation	Input	Practice
	Control group				
Achievement improvement on remembering	Static	--	.07	.27	.26
	Simple animation	--	--	.20	.18
	Input	--	--	--	-.02
	Practice	--	--	--	--
Achievement improvement on understanding	Static	--	.81	.69	.80
	Simple animation	--	--	-.1	.01
	Input	--	--	--	.11
	Practice	--	--	--	--
Achievement improvement on lower-level applying	Static	--	-.04	.55	.46
	Simple animation	--	--	.57	.48
	Input	--	--	--	-.09
	Practice	--	--	--	--
Achievement improvement on higher-level applying	Static	--	.31	.26	.47
	Simple animation	--	--	-.06	.20
	Input	--	--	--	.25
	Practice	--	--	--	--
Confidence improvement on remembering	Static	--	.59	.45	.48
	Simple animation	--	--	-.16	-.15
	Input	--	--	--	.02
	Practice	--	--	--	--
Confidence improvement on understanding	Static	--	.34	.30	.46
	Simple animation	--	--	-.05	.01
	Input	--	--	--	.14
	Practice	--	--	--	--
Confidence improvement on lower-level applying	Static	--	.38	.36	.13
	Simple animation	--	--	-.02	-.26
	Input	--	--	--	-.24
	Practice	--	--	--	--
Confidence improvement on higher-level applying	Static	--	.44	.38	.53
	Simple animation	--	--	-.07	.00
	Input	--	--	--	.08
	Practice	--	--	--	--
Program perception	Static	--	.11	.12	.47
	Simple animation	--	--	.00	.32
	Input	--	--	--	.35
	Practice	--	--	--	--

References

Albert, J. H. (1993). Teaching Bayesian statistics using sampling methods and MINITAB. *The American Statistician*, *47*, 182-191.

Alesandrini, K. L. (1984). Pictures and adult learning. *Instructional Science, 13*, 63–77.

Allen, M. P. (2007). Educational aspects of molecular simulation. *Molecular Physics, 105*, 157-166.

Anderson, L. W. & Krathwohl, D. R. (Eds.). (2001). *A taxonomy for learning, teaching, and assessing: A revision of Bloom's taxonomy of educational objectives*. New York: Longman.

Anglin, G. J., Vaez, H., & Cunningham, K. L. (2004). Visual representations and learning: The role of static and animated graphics. In D. H. Jonassen (Ed.), *Handbook of research for educational communications and technology* (2nd ed., pp. 865–916). Mahwah, NJ: Lawrence Erlbaum Associates.

Atkinson, R. & Shiffrin, R. (1968). Human memory: A proposed system and its control processes. In K Spence & J Spence (Eds.). *The psychology of learning and motivation: Advances in research and theory* (Vol. 2). New York: Academic Press.

Bandura, A. (1997). *Self efficacy: The exercise of control.* NY: Freeman.

Bangert-Drowns, R. L., Kulik, C. C., Kulik, J. A., & Morgan, M. (1991). The instructional effect of feedback in test-like events. *Review of Educational Research, 61,* 213-238.

Ben-Zvi, D. (2000). Toward understanding the role of technological tools in statistical learning. *Mathematical Thinking and Learning, 2* , 127-155.

Betrancourt, M. (2005). The animation and interactivity principles in multimedia learning. In Mayer (Ed.), *The Cambridge handbook of multimedia learning* (pp.287 -296). New York: Cambridge.

Bloom B. S. (1956). *Taxonomy of Educational Objectives, Handbook I: The Cognitive Domain.* New York: David McKay.

Boucheix, J. & Schneider, E. (2008). Static and animated presentations in learning dynamic mechanical systems. *Learning and Instruction*. Retrieved from http://www.sciencedirect.com/science/article/B6VFW-4S8TW2S-2/2/40057b78cb2557f941e425b66efe7c95

Bryce, G. R. (2005). Developing tomorrow's statisticians, *Journal of Statistics Education, 13*(1). Retrieved from http://www.amstat.org/publications/jse/v13n1/bryce.html

Chance, B., Ben-Zvi, D., Garfield, J., & Medina, E. (2007). The role of technology in improving student learning of statistics. *Technology Innovation in Statistics Education, 1*(1), 1-25.

Curriculum guidelines for undergraduate programs in statistical science. (2010) . Retrieved Feb 2, 2010, from http://www.amstat.org/education/curriculumguidelines.cfm

Dambolena, I. G. (1986). The t distribution and computer simulation. *Mathematics and Computer Education, 20*(1), 10-13.

Dambolena, I. G. (1986). Using Simulation in Statistics Courses. *Collegiate Microcomputer, 4*, 339-344.

DelMas, R. C., Garfielf, J., & Chance, B. L. (1999). A model of classroom research in action: developing simulation activities to improve students' statistical reasoning. *Journal of Statistics Education, 7*(3). Retrieved from www.amstat.org/publications/jse/secure/v7n3/delmas.cfm

Dempsey, J. V., Driscoll, M. P., & Swindell, L. (1993). Text-based feedback. In J.V. Dempsey & G.C. Sales (Eds.), *Interactive instruction and feedback* (pp. 21-53). Englewood Cliffs, NJ: Educational Technology Publications.

Dinov, I. D., Sanchez, J., & Christou, N. (2008). Pedagogical utilization and assessment of the statistic online computational resource in introductory probability and statistics courses. *Journal of Computers & Education*, *50*, 284–300.

Draves, W. A. (2000). *Teaching Online*. River Falls, WI: LERN Books.

Duchastel, P. C. (1978). Illustrating instructional texts. *Educational Technology, 11*, 36-39.

Evans, C. & Gibbons, N. (2007). The interactivity effect in multimedia learning, *Computers & Education, 49*, 1147–1160.

Evans, C. & Sabry, K. (2002). Evaluation of the interactivity of Web-Based learning systems: principles and process. *Journal of Innovations in Education and Teaching International, 40* (1), 89-99.

Falk, R. (1986). Misconceptions of statistical significance. *Journal of Structural Learning, 9*, 83-96.

Gal, I. & Garfield, J. (1999). Assessment and statistics education: current challenges and directions. *International Statistical Review, 67(*1), 1-12.

Gao, T. & Lehman, J. D. (2003). The effects of different levels of interaction on the achievement and motivational perceptions of college students in a Web-based learning environment. *Journal of Interactive Learning Research*, *14*, 367-386.

Garfield, J. (1995). How students learn statistics. *International Statistics Review, 63*(1), 25-34.

Garfield, J. (2002). The challenge of developing statistical reasoning. *Journal of Statistics Education, 10*(2). Retrieved from http://www.amstat.org/publications/jse/v10n3/garfield.html

Garfield, J. & Ahlgren, A. (1988). Difficulties in learning basic concepts in probability and statistics: implications for research. *Journal for Research in Mathematics Education, 19*(1). 44-63.

Gordon, F. (1987). Computer Graphics Simulation of the Central Limit Theorem. *Mathematics and Computer Education, 21*(1), 48-55.

Groeneboom, P., Jong, P., Tischenko, D.B., & Zomeren, B. (1996). Computer-assisted statistics education at Delft University. *Journal of Computational & Graphical Statistics, 5*, 386-399.

Hegarty, M. & Just, M. A. (1989). Understanding machines from text and diagrams. In H. Mandl & J. R. Levin (Eds.), *Knowledge acquisition from text and pictures* (pp. 171–194). Amsterdam: Elsevier.

Hubbard, R. (1992). Teaching statistics with Minitab. *The Australian Mathematics Teacher, 48*(4), 8-10.

Kamthan, P. (1999). *JAVA applets in education.* Retrieved from http://www.tech.irt.org/articles/js151

Knapp, T. R. & Schafer, W. D. (2009). From Gain Score t to ANCOVA F (and vice versa). *Practical Assessment, Research & Evaluation, 14*(6), 1-7.

KOC, M. (2005). Individual learner differences in web-based learning environments: from cognitive, affective and social-cultural perspectives. *Turkish Online Journal of Distance Education, 6*(4), 12-22.

Kopcha, T. J. & Sullivan, H. (2008). Learner preferences and prior knowledge in learner-controlled computer-based instruction. *Educational Technology Research and Development, 56*, 265-286.

Kristof, R. & Satran, A. (1995). *Interactivity by design: creating and communicating with new media.* Mountain View, CA: Adobe Press.

Kulhavy, R. W. & Stock, W. A. (1989). Feedback in written instruction: The place of response certitude. *Educational Psychology Review, 1*(4), 279-308.

Kulhavy, R. W. & Wager, W. (1993). Feedback in programmed instruction: Historical context and implications for practice. In J. V. Dempsey & G. C. Sales (Eds.), *Interactive instruction and feedback* (pp. 3–20). Englewood Cliffs, NJ: Educational Technology.

Kulik, J. A. (2002). School Mathematics and Science programs benefit from instructional technology. Retrieved from http://www.nsf.gov/sbe/srs/infbrief/nsf03301/start.htm

Landewehr, J., Swift, J., & Watkins, A. E. (1987). *Methods of Sampling. In Exploring Surveys and Information from Samples*. Palo Aldo, CA: Dale Seymour.

Landewehr, J., Watkins, A. E., & Swift, J. (1987). E*xploring surveys: Information from samples*. Palo Alto, CA: Dale Seymour.

Levie, W. H. & Lentz, R. (1982). Effects of text illustrations: A review of research. *Educational Communication and Technology Journal, 30*, 195–232.

Levin, J. R. (1981). On the functions of pictures in prose. In F. J. Pirozzolo & M. C. Wittrock (Eds.), *Neuropsychological and cognitive processes in reading* (pp. 203-228). New York: Academic Press.

Litchfield, B. (1993, April). Design factors in multimedia environments: Research findings and implications for instructional design. Paper presented at the annual meeting of the American educational research association (AERA), Atlanta, GA.

Lock, R. H. (1983). Using computer-simulation in teaching statistics. *American Mathematical Monthly, 90*(10), 114.

Lowe, R. (2004). Interrogation of dynamic visualization during learning. *Instruction and Learning, 14*, 257-274.

Marasinghe, M. G., Meeker, W. Q., Cook, D., & Shin, T. (1996). Using graphics and simulation to teach statistical concepts. *The American Statistician, 50*, 342-351.

Mayer, R. E. (2009). *Multimedia learning (2nd ed.)*. New York: Cambridge University Press.

Mayer, R. E. & Anderson, R. B. (1992). The instructive animation: Helping students build connections between words and pictures in multimedia learning. *Journal of Educational Psychology, 84*, 444–452.

Mayer, R. E. & Chandler, P. (2001). When learning is just a click away: does simple user interaction foster deeper understanding of multimedia messages? *Journal of Educational Psychology, 93*, 390-397.

Mayer, R. E. & Gallini, J. K. (1990). When is an illustration worth ten thousand words? *Journal of Educational Psychology, 82*, 715-726.

Mayer, R. E. & Moreno, R. (2002). Animation as an aid to multimedia learning. *Educational Psychology Review, 14*(1), 87–99.

Mittag, K. C. & Taylor, S. (1996, April). *Using graphing calculator technology in educational statistics courses*. Paper presented at Annual Meeting of American Educational Research Association (AERA), New York, NY.

Moore, D. (1997). New pedagogy and new content: the case of statistics. *International Statistics Review, 65*, 123-165.

Moore, M. G. (1989). Three types of interaction. *The American Journal of Distance Education, 3* (2), 1-6.

Mory, E. H. (2003). Feedback research revisited. In D.H. Jonassen (Ed.), *Handbook of research for educational communications and technology* (pp. 745-783). New York: MacMillan Library Reference.

Naps, T. L., Rößling, G., Almstrum, V., Dann, W., Fleischer, R., Hundhausen, C., Korhonen, A., Malmi, L., McNally, M., Rodger, S., & Velazquez-Iturbide, J. Á. (2003). Exploring the role of visualization and engagement in computer science education, *SIGCSE Bulletin, 35*, 2, 131-152.

National Council of Teachers of Mathematics. (1989). *Curriculum and Evaluation Standards For School Mathematics*. Reston, VA.

Paivio, A. (1991). Dual coding theory: Retrospect and current status. *Canadian Journal of Psychology, 45*, 255–287.

Park, O. & Hopkins, R. (1993). Instructional conditions for using animated visual displays: A review. *Instructional Science, 22*, 1–24.

Phye, G. D. & Bender, T. (1989). Feedback complexity and practice: Response pattern analysis in retention and transfer. *Contemporary Educational Psychology, 14,* 97-110.

Rieber, L. P. (1990). Animation in computer-based instruction. *Educational Technology Research and Development, 38*(1), 77–86.

Rieber, L. P. (1991). Effects of visual grouping strategies of computer-animated presentations on selective attention in science. *Educational Technology Research and Development, 39*(4), 5-15.

Rosebery, A. S. & Rubin, A. (1989). Reasoning under uncertainty: developing statistical reasoning. *Journal of Mathematical Behavior*, *8*, 205-219.

Rosebery, A. S. & Rubin, A. (1990). Teaching statistical reasoning with computers. *Teaching Statistics, 12*(2), 38-43.

Savoldelli, G., Naik, V., Hamstra S., & Morgan, P. (2005). Barriers to use of simulation-based education, *Can J Anaesth*, *52*, 944–950.

Schau, C., Stevens, J., Dauphinee, T. L., & Del Vecchio, A. (1995). The development and validation of the Survey of Attitudes Toward Statistics. *Educational and Psychological Measurement*, *55*, 868-875.

Schnotz, W. & Grzondziel, H. (1999). Individual and co-operative learning with interactive animated pictures. *European Journal of Psychology of Education, 14*, 245-265.

Schwier, R. A. & Misanchuk, E. (1993). *Interactive multimedia instruction.* Englewood Cliffs, NJ: Educational Technology Publications.

Segenchuk, S. (1997). *The Role of visualization in education*. Retrieved from http://web.cs.wpi.edu/~matt/courses/cs563/talks/education/IEindex.html

Sims, R. (1997). Interactivity: A Forgotten Art? *Computers in Human Behavior, 13*, 157-180.

Smith, B. (2003). Using and evaluating resampling simulations in SPSS and Excel. *Teaching Sociology*, *31*, 276-287.

Smith, G. (1998). Learning statistics by doing statistics. *Journal of Statistics Education*, *6*(3). Retrieved from http://www.amstat.org/publications/jse/v6n3/smith.html

Somyurek, S., Guyer, T., & Atasoy, B. (2008). The effects of individual differences on learner's navigation in a courseware. *The Turkish Online Journal of Educational Technology, 7*(2), 32-40.

Spaulding, J. J. (2007). *Multimedia-based classroom instruction and student achievement: What is the relationship between multimedia-based classroom*

instruction and student achievement in a public elementary school district. Unpublished doctoral dissertation, University of La Verne. California.

Spiro, R. J., Feltovich, P. J., Jacobson, M. J., & Coulson, R. L. (1992). Cognitive flexibility, constructivism, and hypertext: Random access instruction for advanced knowledge acquisition in ill-structured domains. In T. M. Duffy & D. H. Jonassen (Eds.), *Constructivism and the technology of instruction: A conversation* (pp. 57-76). Hillsdale, NJ: Lawerence Erlbaum Associates.

Spiro, R. J. & Jehng, J. C. (1990). Cognitive flexibility and hypertext: Theory and technology for the nonlinear and multidimensional traversal of complex subject matter. In D. Nix & R. Spiro (Eds.), *Cognition, education, and multimedia: Exploring ideas in high technology* (pp. 163-205). Hillsdale, NJ: Lawrence Erlbaum Associates.

Steen, L. A. (1999). *Projects in Quantitative Literacy.* Retrieved from http://www.amstat.org/publications/jse/v6n3/smith.html

Structural Informatics Group. (2007). *Digital Anatomist Project.* Retrieved from http://www9.biostr.washington.edu/da.html

Teo, H., Oh, L., Liu, C., & Wei, K. (2003). An empirical study of the effects of interactivity in Web user attitude. *International Journal of Human-Computer Studies, 58*, 281-305.

Tversky, B., Morrison, J. B., & Betrancourt, M. (2002). Animation: Can it facilitate? *International Journal of Human–Computer Studies, 57*, 247–262.

Vaughn, B. (2009). *The language of statistical decision making.* Unpublished booklet for classes, University of Texas at Austin, Texas.

Vekiri, I. (2002). What is the value of graphical displays in learning? *Educational Psychology Review*, *14*, 261–312.

Velleman, P. F. & Moore, D. S. (1996). Multimedia for teaching statistics: promises and faults. *The American Statistician*, *50*, 217-225.

Williams, F., Rice, R. E., & Rogers, E. M. (1989). *Research methods and the new media.* New York: The Free Press.

Windsor, N. J. (1998). Simulating a sampling problem using a T1-83 graphics calculator. *Australian Senior Mathematics Journal, 12*(2), 32-37.

Witten, I. H. & Frank, E. (2005). *Data mining: Practical machine learning tools and techniques*. San Francisco, CA: Morgan Kaufmann.

Vita - Pei-Yu Wang

Pei-Yu Wang was born in 1979 in Taichung, Taiwan. In 2002, she received her B.Ed. and B.B.A. in Education and Management Information Systems (double major) from National Chengchi University. In 2004, she received her M.S. in Information and Computer Education from National Taiwan Normal University. She started her doctoral studies at UT Austin in 2005. She is interested in the design and development of online systems, including online knowledge sharing systems, open-source Web systems, and instructional multimedia systems. During her graduate career, she participated in many research projects. She worked with Dr. Janet Mei-Chuen Lin to create the online knowledge sharing systems (funded by National Science Council in Taiwan) from 2002 to 2005. Then she participated in Dr. Min Liu's instructional multimedia projects (e.g., Not Just Sushi and Alien Rescue) from 2005 to 2009. In addition, she helped college professors to create online instructional systems and multimedia programs for seminar courses, statistics courses, and teachers' workshops. After graduation, Pei-Yu hopes to go back Taiwan to contribute to educational innovation and reformation by introducing new educational technologies and methodologies.

Permanent address/ e-mail address: sweetpeiyu@gmail.com

This dissertation was typed by Pei-Yu Wang.