The Dissertation Committee for Michael Charles Mayrath
certifies that this is the approved version of the following dissertation:

Examining Factors that Affect Performance in Complex Simulation Environments

Committee:

Daniel H. Robinson, Supervisor

Marilla D. Svinicki

Tiffany A. Whittaker

Brandon K. Vaughn

Paul E. Resta

Examining Factors that Affect Performance in Complex Simulation Environments

by

Michael Charles Mayrath, B.J.; M.A.

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 2009

Dedication

To my grandmother, Jimmy, who is the nucleus of our family and our model of love.

Acknowledgements

I would like to thank my dissertation committee for their wisdom and guidance: Dr. Dan Robinson, Dr. Paul Resta, Dr. Marilla Svinicki, Dr. Tiffany Whittaker, and Dr. Brandon Vaughn. For over five years, my supervisor, Dan Robinson, has provided me with advice, opportunities, and mentorship that have opened a new world of possibilities for me, and for that, I will always be grateful to him. I know we will collaborate for many years to come.

I would like to thank people all over the campus of the University of Texas at Austin for their help with my dissertation. Thanks to my fellow Area I colleagues for helping to conceptualize, pilot test, and edit my dissertation, especially Priya Nihalani, Shana Shaw, Camilo Guerrero, Robert Ellis, Frank Roberts, and Christina Cestone. Thanks to everyone at the Learning Technology Center, including Rob Donald, Nancy Bell, James Keys, and Chris Yallalee. Without their help setting up the lab and scheduling meetings this study would not have been completed. Thanks to my colleagues at the Institute for Advanced Technology for allowing flexibility in my work schedule and presenting constantly challenges problems to solve, in particular Sheilagh O'Hare, Aubrey White, John Gentle, General Butch Funk, and the rest of the University XXI team.

And, thanks to all my friends and family for their patience over the past five years at times when I had to put "work" first.

Examining Factors that Affect Performance in Complex Simulation Environments

Publication No._____________

Michael Charles Mayrath, Ph.D.
The University of Texas at Austin, 2009

Supervisor: Daniel H. Robinson

This study examined the effects of manipulating the modality (text-only, voice-only, voice+text) of a tutorial and restriction (restricted vs. unrestricted) of a simulation's interface on retention and transfer of tutorial content. The tutorial prepared novice students to use Packet Tracer, a simulation developed by Cisco that teaches network engineers how to build and troubleshoot computer networks. Retention was measured using a multiple choice test whereas transfer was measured using an assessment embedded within Packet Tracer.

An interaction was found between modality and restriction on the Packet Tracer transfer test. When Packet Tracer's interface was unrestricted, students who received the voice-only tutorial performed significantly better on the transfer test than students who received the text-only tutorial. This finding is consistent with the cognitive theory of multimedia learning and previous research on modality effect. However, this is also an original finding because previous research has not examined the interaction between a tutorial's modality and the restriction of a complex simulation's interface. This study addressed relevant instructional technology design questions, such as how to design tutorials for

complex simulations and what effect restricting a simulation's interface has on retention and transfer for novice students.

Table of Contents

List of Tables ... X

List of Figures ... XI

Chapter 1 Introduction ... 1

- Background ... 1
- Two Areas Needing Research ... 4
- Statement of problem ... 7
- Purpose of study ... 7
- Key variables ... 9
- Research questions and Hypotheses ... 9

Chapter 2 Review of Literature ... 11

- Simulations ... 11
- Packet Tracer ... 13
- Sims and Learning Theories ... 18
- Cognitive Load ... 20
- Cognitive Theory of Multimedia Learning ... 30
- Assessment ... 33

Chapter 3 Methods ... 36

- Participants and Design ... 36
- Materials and Apparatus ... 37
- Procedure ... 44

Chapter 4 Results ... 48

- Scoring Issue ... 48
- Overview ... 51
- Pretest ... 52
- Retention Test ... 54
- Transfer Test ... 60
- Self-Report Data ... 66

Chapter 5 Discussion ... 79

- Summary ... 79
- Findings ... 82

Research Questions 83
Lack of Findings for Cognitive Load........ 88
Lost Data........ 89
Challenges of Educational Technology Research........ 90
Recommendations for Running Educational Technology Research in a Computer Lab..... 92
Implications 94
Limitations........ 95
Recommendations for future research 97
Conclusion........ 99

Appendices
Appendix A – Overview of the cognitive theory of multimedia learning........ 101
Appendix B – Pretest of computer networking knowledge 102
Appendix C – Retention post-test........ 104
Appendix D – Packet Tracer survey 106
Appendix E – Mental effort rating........ 108
Appendix F – Difficulty rating. 108
Appendix G – Task analysis for tutorial and PT activity........ 109
Appendix H – Instructions........ 110
Appendix I – Consent form 111
Appendix J – Certificate of completion........ 113
Appendix K – Results from full sample 114
Appendix L – Scatter plots for each group 117

References 120

Vita …........ 124

List of Tables

Table 3.1 – Experiment Design 37

Table 3.2 – Assessments, Measures, & Questionnaires 40

Table 3.3 – Data collection sessions 45

Table 4.1 – Mean and Standard Deviations for All Groups 51

Table 4.2 Kolmogorov-Smirnov Z for the Pretest by Group 53

Table 4.3 Kolmogorov-Smirnov Z for the Retention test by Group 56

Table 4.4 Kolmogorov-Smirnov Z for the PT Activity by Group 61

Table 4.5 – ANOVA Source Table: PT activity measure 62

Table 4.6 – Means for Two-Way Interaction on PT Activity 63

List of Figures

Figure 2.1 – Packet Tracer's User Interface 15

Figure 2.2 – Packet Tracer's Assessment Tree 16

Figure 2.3 – Packet Tracer's Virtual Representation of the front of a PC 17

Figure 2.4 – Information Processing Model (Mayer, 2001) 22

Figure 3.1 – Screenshot of the Text-only Tutorial 38

Figure 3.2 – PT Restricted: locked parts of interface shown in red 39

Figure 3.3 – PT's instructions 39

Figure 3.4 – Study Session Procedure 46

Figure 4.1 – Histogram for Pretest 53

Figure 4.2 – Histogram for Retention post-test 56

Figure 4.3 – Scatter plot for the pretest and retention post-test 58

Figure 4.4 – Histogram for PT Activity 61

Figure 4.5 – Two-Way Interaction on PT Activity 64

Figure 4.6 – Results for Self-Report item "PT's interface was easy to use." 67

Figure 4.7 – Results for Self-Report item "PT was frustrating." 68

Figure 4.8 – Results for Self-Report item "Packet Tracer was fun." 69

Figure 4.9 – Results for Self-Report item "Packet Tracer's interface was overwhelming." 69

Figure 4.10 – Results for Self-Report item "Locked messages distracted me." 70

Figure 4.11 – Results for Self-Report item "Not being able to access restricted parts of Packet Tracer made me curious." 71

Figure 4.12 – Results for Self-Report item "Mental Effort during tutorial." 72

Figure 4.13 – Results for Self-Report item "Mental Effort using PT." 72

Figure 4.14 – Results for Self-Report item "Tutorial's level of complexity." 73

Figure 4.15 – Results for Self-Report item “Mental Effort using PT.” 73

Figure 4.16 – Results for Self-Report item “Difficulty understanding the tutorial.”...................... 74

Figure 4.17 – Results for Self-Report item “Difficulty remembering information
from the tutorial.” 74

Figure 4.18 – Results for Self-Report item “Difficulty using Packet Tracer.” 75

Figure 4.19 – Results for Self-Report item “Eagerness to continue using Packet Tracer.”............ 76

Figure 4.20 – Results for Self-Report item “How interesting is this material?”............................ 76

Figure 4.21 – Results for Self-Report item “How much effort did you put into this study?”......... 77

Chapter 1 Introduction

Background

Leaders' visions

At recent *Technology, Entertainment, & Design* (TED Talks) conferences, Joy (2006) and Negroponte (2007) presented their visions for the future of educational technology. Joy co-founded Sun Microsystems, sits on the boards of top corporations, and controls billions of dollars in venture capital for technology development. Negroponte leads the Massachusetts Institute of Technology's Media Lab and founded the *One Laptop Per Child* (OLPC) program. Negroponte stated that OLPC has the goal of providing all students around the world with inexpensive laptops and that steady progress is being made towards that goal. Joy discussed Moore's Law, which describes technological change and states that computer processing speed doubles every two years with this trend continuing through 2029 (Gelsinger, 2008; Moore, 2003). Based on Moore's Law and the goals of OLPC, Joy foresaw that in 10 years children around the world will have low cost, wirelessly connected laptops as powerful as today's most expensive desktop computer.

Joy (2006) challenged educators and technologists to develop the kinds of educational tools that can take advantage of these powerful and inexpensive computers. He stated that "today we have very good computers, but we don't have very good software for them." Powerful, cost efficient laptops that wirelessly connect to the internet can facilitate international communities of practice and collaborative learning environments that will make the Pen Pal letter system of yesteryear look like the ancient past. However, Joy cautioned that technology alone will not increase meaningful learning. His foresight is evidenced by public schools abandoning one-to-one student to laptop programs because laptops became a source of distraction, cheating, and frequently broke down. Further, teachers often do not know how technology pedagogically fits into the class (Ertmer, 2005).

Recently, the U.S. Department of Education (2008) published a report titled, *Harnessing Innovation to Support Student Success: Using Technology to Personalize Education*, which was based on 18 months of research and conclusions from four roundtable discussions with experts ranging from CEOs to teachers and students. The report stated that while most classrooms around the country are wired, there is no evidence that it has had a transformative effect on education. Consistent with Joy's (2006) comments, this report stated that the computers and connectivity are in place; however, schools and teachers have been unable to produce a large-scale, measurable impact from technology (Cuban, 2001; Ertmer, 2005; Zhao & Frank, 2003). The underlying assumption here is that technology *can* improve education. Yet, technology is just a tool and therefore can be used either effectively or ineffectively.

When used effectively, technology has been found to significantly increase learning, enhance student motivation and engagement, provide opportunities for social interaction, facilitate distance collaboration, and allow for exploration, experimentation, creation, and interaction within a rich environment (Atkinson & Wilson, 1968; Bransford, 2003; Dede, 1995; Dede, Clarke, Ketelhut, Nelson, & Bowman, 2005; Gee, 2003; Mayer, 2005; Moreno, Mayer, Spires, & Lester, 2001; Papert, 1980; Resta & Laferrière, 2007).

However, when used ineffectively, technology is distracting, frustrating, and can significantly decrease learning (Clark & Choi, 2005). There are no set standards for educational technology because learning is situationally-based in terms of context variables, such as the learner, content, task, assessment, and feedback. Technology adds another variable to this already complex environment and usually requires professional development for the instructor and technical support. Yet, the research field of educational technology has only existed for a few decades (Atkinson & Wilson, 1968; Papert, 1980), and given the exponential rate of technology's evolution over the last 10 years, the ways in which technology can be utilized for education are constantly changing. What we do know is that there is great potential for highly interactive, social learning experiences and increased possibilities for assessment.

Simulations

Simulations (sims) are one way technology can be used effectively in education (Rieber, 2005). Sims allow learners to apply and test the limitations of their knowledge, skills, and abilities (Gredler, 2004; Jonassen & Rohrer-Murphy, 1999; O'Neil, Wainess, & Baker, 2005). The sim examined in this study is what Gredler refers to as a *symbolic systematic simulation*, which is a sim whereby the student "tests his or her conceptual model of the relationships among the variables in the system" (p. 574). Individuals use symbolic systematic sims to practice carrying out complex processes, such as analyzing, diagnosing, and correcting problems. For example, flight simulators give pilots practice flying a plane and preparing for worst case scenarios. Medical schools use virtual sims for students to practice surgery and high-stakes decision making without life and death consequences. Additionally, at the Cisco Networking Academy, computer networking sims are used by thousands of students around the world to practice creating and troubleshooting networks. This study used a sim developed by Cisco called *Packet*

Tracer (PT), which is used worldwide to train engineers how to build and troubleshoot computer networks.

Sim Challenges

There are numerous challenges associated with using complex sims. First, they generally require training on how to use them before a student can utilize a sim to its full pedagogical potential. Thus, tutorials are often made to accompany complex sims; however, a tutorial's effectiveness is determined by how well it is designed, and numerous tutorials are poorly designed because developers are unfamiliar with principles of multimedia learning design (Mayer, 2001, 2005). Second, novices and experts have very different schemas of domain knowledge and have different cognitive load capacity limits when operating within that domain (Kalyuga, Ayres, Chandler, & Sweller, 2003). Thus, should complex sims be designed differently for experts and novices? Or, are there other methods to not overloading novice students' load capacities while using an expert level sim interface? Again, sim developers frequently do not take cognitive load limitations into consideration when designing a sim user interface.

Two Areas Needing Research

This study examines two distinct areas of research needing to be explored to address the challenges facing sims. This research informs our understanding of how to use complex sims for learning and assessment.

Tutorials & Modality of Instruction

First, sims such as PT are complex and can frustrate users when no training is provided on how to use the sim. *My First PT Labs* are a series of tutorials that teach Cisco students how to use PT appropriately. The tutorials are 3-6 minute animated

Captivate screencasts that orient the user to PT's interface by showing how to drag network icons and symbols to the sim's workspace. The narration in the tutorial is delivered through pop-up text-boxes that accompany the animation. There is no voice narration. Given recent research, there may be more effective ways to present the information in these tutorials.

According to the cognitive theory of multimedia learning, specifically the modality effect and the redundancy principle, the current text and animation tutorials cause students to split their attention by having to read the text boxes while also watching the cursor in the animation (Mayer, 2005; Sweller, 2005). Further, why only use the visual channel to deliver instruction, why not use the verbal channel also? Research has consistently shown that an animation accompanied by audio narration is more effective on retention and transfer measures when compared to an animation with text boxes for narration (Ginns, 2005; Mayer, 2001). However, researchers have found that an audio narration accompanied by short pieces of text can be effective in directing attention to the critical parts of an animation (Moreno & Mayer, 2002). While a little text may help, researchers still assert that providing a full text of the narration at the same time as an animation can result in ineffective learning outcomes (Mayer & Johnson, 2008).

As research over the last 10 years has shown, the *modalities* in which information is presented in multimedia learning environments can be a testable variable that informs both theory and practice. While the *modality* (text-only, voice-only, and voice+text) of instruction has been shown to affect retention and transfer outcomes (Mayer, 2001, 2002; Mayer & Johnson, 2008; Moreno, 2006), these studies have examined instructional contexts, such as explanations of lighting, brakes, science, and, botany. However, there has been no research testing the modality effect in a tutorial that teaches novice students how to use a complex simulation. What is the best instructional delivery modality for such tutorials?

Simulation Interface & Cognitive Overload

Second, whether novice or expert in domain knowledge, it can take time to learn how to use a complex sim's interface or visual display. Experts are at an advantage, however, because they have a greater cognitive load capacity when working within that domain compared to novices due to the expert's well formed schemas (Kalyuga et al., 2003; Sweller, 1988). For example, if a person with no prior experience had to land a plane, all the buttons, switches, and levers on the control panel would be overwhelming. What if the available options were reduced to only the controls needed to land the plane? This form of scaffold might help minimize the options for error or distraction. In a sim, restricting access to extraneous features and functions can reduce a novice user's potential for error. Thus, a sim should present novices with less information and stimuli; yet, sims should also provide experts with advanced tools for hypothesis testing and exploration.

Building novice and expert versions of the same sim is impractical, and designing two interfaces for the same sim requires extensive programming. One solution is to simply *restrict* or take away the functionality of extraneous elements within a novice user's interface. The rationale being that if novice students only have access to the essential parts of the sim needed for completing a task, then they will not waste their time and effort on extraneous sim functionality. This solution is relatively simple in terms of programming and therefore cost effective. However, there is no research on what effect restricting access to parts of a sim's interface has on novice sim user behavior and on their performance accomplishing tasks with the sim.

Statement of problem

Research is needed to develop and tie together two inter-related topics of educational technology research that can impact the learner: (1) Instructional Modality and (2) Simulation Interface. First, while there is research on the effect of modality in numerous types of multimedia learning environments (Mayer, 2001, 2005; Sweller, 2005), no research currently exists on modality and the tutorials that train how to use a complex sim.

Second, sim developers do not produce novice and experts versions due to costs. However, based on cognitive load theory, restricting access to the extraneous parts of a sim's interface may reduce extraneous cognitive load for novices because unneeded functionalities would be inaccessible. I was unable to locate any research on what effect restricting access to certain parts of a sim's interface has on novice sim user behavior and performance on the task.

Purpose of study

The purpose of the current study was to inform sim developers by examining design variables that affect student learning outcomes of retention and transfer. The findings from this study have implications for sim designers and instructional designers. Knowing how modality and restricting access to parts of a sim's interface affect cognitive load can lead to more effective educational software and improved learning.

Study Overview

This study examined the effect of manipulating tutorial modality (text-only, voice-only, voice+text) and sim interface restriction (restricted vs. unrestricted) on the dependent measures of retention and transfer. The overarching research questions were what are the optimal modalities for presenting a tutorial on how to use PT, and what effect does restricting access to extraneous parts of PT's interface have on the dependent measures. Researchers evaluating learning in multimedia environments have traditionally used measures of retention and transfer (Ginns, 2005; Mayer, 2001, 2005). Thus, the rationale for using a retention test and a transfer test is based on previous modality research. In this study, I measured subjects' ability to retain information from the tutorial and their ability to transfer the knowledge from the tutorial to PT. While using PT, subjects' were assessed on their ability to troubleshoot a network. PT's internal assessment tree scored the number of correct troubleshooting solutions for each subject, which was considered a measure of transfer.

Subjects first took a 10-item pretest of computer networking knowledge. Next, they watched a 10-minute tutorial on computer networking concepts and how to use PT to create, setup, and test a simple network connection between a computer and server. Following the tutorial subjects used PT for the *Troubleshooting* activity. Finally, subjects completed a 14-item post-test of retention from the tutorial and a self-report survey asking them to rate various aspects of their experience.

The sample used in this study consisted of subjects from the University of Texas at Austin College of Education subject pool. Due to a scoring issue explained later in depth, 81 scores were used for the retention test. Eleven PT activity transfer test scores were lost due to crashes and other issues; thus, 70 subjects were used for the PT activity transfer test. An analysis of covariance (ANCOVA) was used to test for an interaction and main effects on the retention test. A computer networking knowledge pretest correlated with the retention post-test but not the PT activity transfer test. Thus, the pretest was used as a covariate in the retention ANCOVA but was not used for analysis with the PT activity transfer test, which used an analysis of variance (ANOVA).

Key variables

This study included two between-subjects factors, *tutorial modality* (text-only, voice-only, voice+text) and *sim interface restriction* (restricted vs. unrestricted). Manipulations of modality and interface restriction are examined using two theoretical frameworks: the cognitive theory of multimedia learning (Mayer, 2001) and cognitive load theory (Sweller & Chandler, 1994).

Research questions

RQ1. Given prior research on the modality effect, how does the "My 1st PT Lab" tutorial instructional delivery modality (text-only, voice-only, voice+text) affect performance on dependent measures?

RQ2. What effect does the restriction of PT's interface (restricted vs. unrestricted) have on the dependent measures?

RQ3. Is there an interaction between modality and interface restriction?

Hypotheses

Based on existing literature, it was hypothesized that the high load conditions would overload the subjects' visual channel thus resulting in poorer performance compared to low load conditions (Mayer, 2001; Sweller, 1988).

H1. Research states that voice-only would outperform the other two conditions (Mayer & Johnson, 2008; Moreno & Mayer, 2002).

H2. There is no available literature comparing the effects of a sim's interface being restricted or not. However, it was expected that because these PT users are extreme novices, the restricted group will experience less cognitive overload and therefore outperform the unrestricted group (Kalyuga et al., 2003).

H3. It was predicted that there would not be a significant interaction between modality and interface restriction. It was also predicted that the best condition will be voice-only, restricted.

Summary

The goal of this study was to inform theory and practice by identifying optimal conditions for using, and learning how to use, complex sims. Educational technology as a research field has a good foundation but is still in its infancy in terms of being a science. This study is empirical, grounded in theory, and examines important questions yet to be asked. It contributes to instructional technology research which is important because we in modern, developed nations live in a technology-based society, therefore, we must optimize educational settings through rigorous research. The next Chapter discusses major theories and research relevant to this study.

Chapter 2 Review of Literature

This review begins by defining sims then discussing learning theories related to sims and PT. Next, an in depth review of cognitive load theory is developed to provide a conceptual framework for this study and to explain the latest research on measuring cognitive load. The cognitive theory of multimedia learning is then outlined along with examples of methodologies used in a very productive line of research by Mayer and colleagues. All of these topics are relevant to this study and need explication.

Simulations

According to Gredler (2004), sims have been used as experiential exercises dating back to 16th century war games. Yet, sims are distinguishable from games, an articulation that should be made explicit to provide parameters for what this study is examining. Gredler stated that sims are "open-ended evolving situations with many interacting variables. The goal for all participants is to each take a particular role, address the issues, threats, or problems that arise in the situation, and experience the effects of their decisions" (p. 571). Sims have four defining characteristics: (a) high fidelity representation or model of a complex real-world situation with which a student interacts, (b) roles including responsibilities and constraints that students take on in the sim, (c) an interactive environment where students can explore and test decision making, and (d) feedback for participant actions in the form of changes in the problem or situation.

Sims vs. Games

According to O'Neil and colleagues, there are three major differences between games and sims (O'Neil et al., 2005). First, sims do not generally have competition involved. Rather, they are intended for discovering cause-effect relationships. Authentic cause-effect relationships are at the core of sims and are considered the second major difference. Sims are intended to represent a functioning system, a representation of how something works in the real world (Gredler, 1996, 2004) and are therefore authentic by nature. The third major difference is that with games, the goal structure is linear. In contrast, sims have non-linear goal structures. In a sim, the goal is to create an output through certain actions and decisions. The user manipulates input variables to create output which can then be modified if needed. The process can then be repeated. Thus, the goal structure of a sim is non-linear.

In contrast, games are "competitive exercises in which the objective is to win and players must apply subject matter or other relevant knowledge in an effort to advance in the exercise and win" (Gredler, 2004, p. 571). Games include rules that limit player moves and create game constraints. Incentives or rewards are provided along with penalties for illegal actions. And, rules do not need to be based in reality; whereas, sims are simulated representations of a real-world, causal system (Gredler, 1996, p. 523). Games do not have to represent reality, and they are made to be won. They are also meant to be playful where as sims are used to practice a skill.

Sims are also distinguishable from information delivery types of multimedia learning environments. PowerPoint presentations or animations, such as watching a lesson on how lightning is formed, are passive in comparison to a sim that allows users to interact with the system by controlling variables then witnessing the consequences. A sim on lightning creation would allow the user to control variables that affect lightning creation, such as temperature, electricity, etc. Thus, a well designed sim is expected to provide more opportunities for experiential learning compared to a passive multimedia tutorial (Rieber, 2005). In the present study, both sims and tutorials are examined.

According to Gredler (2004), sim developers must ask themselves two questions. First, "Does the sim meet the criteria for the type of exercise (symbolic or experiential)?" (p. 579). Second, "What is the purpose of the sim? If the sim is to be a culminating experience that involves the application of knowledge, then instruction must ensure that students acquire that knowledge" (p. 579).

Packet Tracer

PT provides a high fidelity, authentic experience that could be classified as both symbolic and experiential. PT was developed to be an experiential learning and assessment tool for the Cisco Networking Academy, which is located in over 160 countries and trains thousands of students every year. PT is integrated into the curriculum so that it is a culminating experience that involves application of knowledge. The knowledge that must be applied is delivered to students through animated tutorials called *My First PT Labs* that show how to build a computer network using PT.

PT provides an authentic representation of what Cisco engineers and network technician must consider when problem solving on the job. The sim is used for group work, homework, formative assessment, hands-on lab reinforcement, lecture demonstrations, modeling and visualization of networking device algorithms and networking protocols, case studies, competitions, and problem-solving activities. PT is used for all levels of Cisco students ranging from novice to expert. However, it should be noted that *novice* Cisco students use PT after first establishing a basic understanding of computer networking.

Frezzo and Stanley (2005) found four problem types to be effective when using PT: *Concept-builders* (model-building that leads to students creating understanding), *Skill-builders* (problem solving that develops procedural knowledge), *Design challenges* (constraint-based problems with multiple correct solutions), *Troubleshooting challenges* (diagnosing, isolating, and fixing the simulated network from a previously bugged network file).

Three pedagogical elements of PT's design create meaningful learning experiences for Cisco students. First, PT allows students to build, configure, and test networks using virtual equipment and connections, see Figure 2.1. Students are able to apply knowledge and skills by problem solving in a virtual environment. Second, PT has an advanced assessment system, *Assessment Tree*, embedded in the program that allows instructors and researchers to create customized activities with automated assessments of the tasks that students must complete to receive credit. The Assessment Tree allows the researcher to check or uncheck whether specific tasks will be assessed when comparing an Initial Model of the network to the Answer Model of the network, see Figure 2.2. This built-in assessment capability allows for custom and detailed measurements to be made.

Third, PT has visual representations (symbols and icons) for all the Cisco products (switches, hubs, routers, etc.) and common computer networking environments (computers, servers, building blueprints, etc.) that a Cisco Network Academy graduate needs to understand. For example, PT has a virtual representation of the front of a PC, see Figure 2.3. A user can turn the PC on or off by clicking on the "Power" button. Similarly, the front and back of the network icons in the sim are exactly the same as in real life. Given these three pedagogical elements of PT, it is reasonable to claim that using PT provides learning experiences that are authentic and situated in the real world context of the tools and technologies Cisco employees will be using on the job. PT therefore provides a good platform for studying how sims can be improved to achieve what Joy and teachers around the country would call “good educational software.”

Figure 2.1 – Packet Tracer’s User Interface

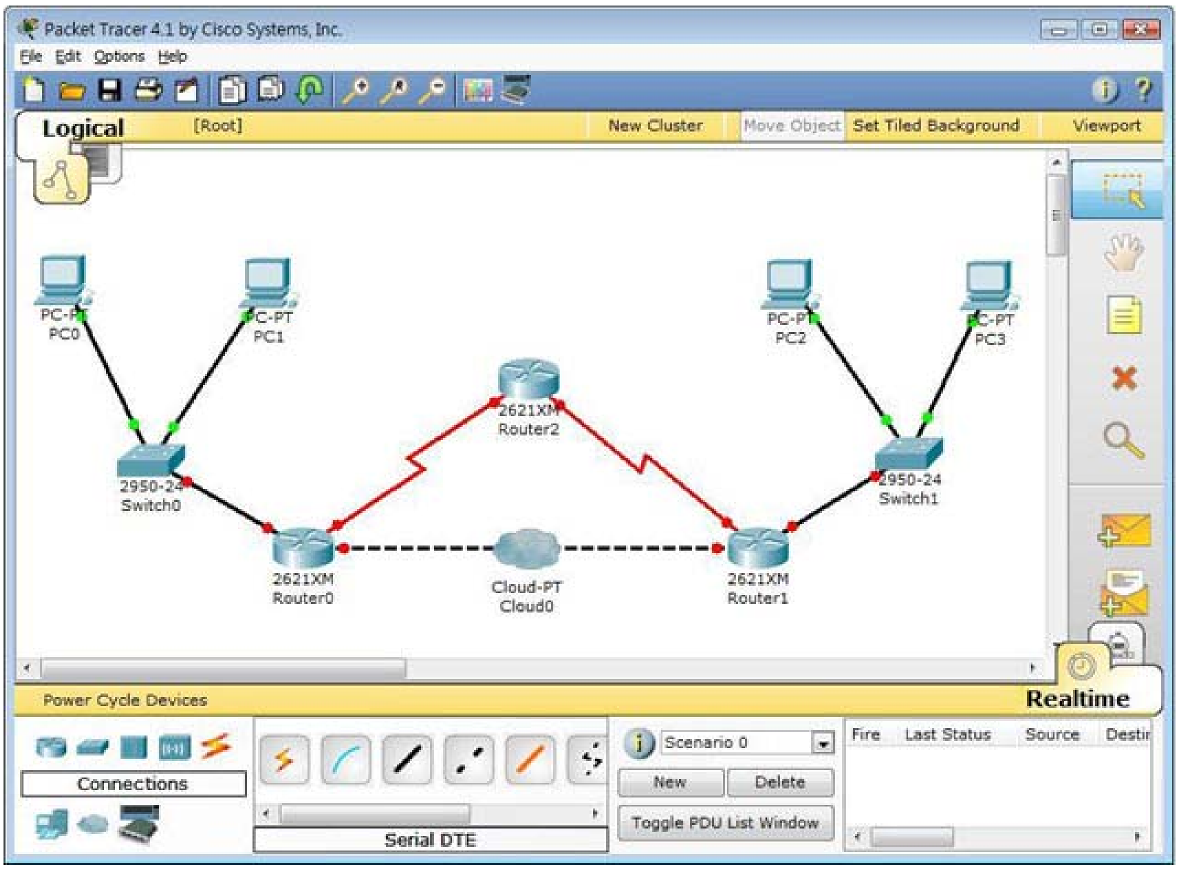


Figure 2.2 – Packet Tracer's Assessment Tree

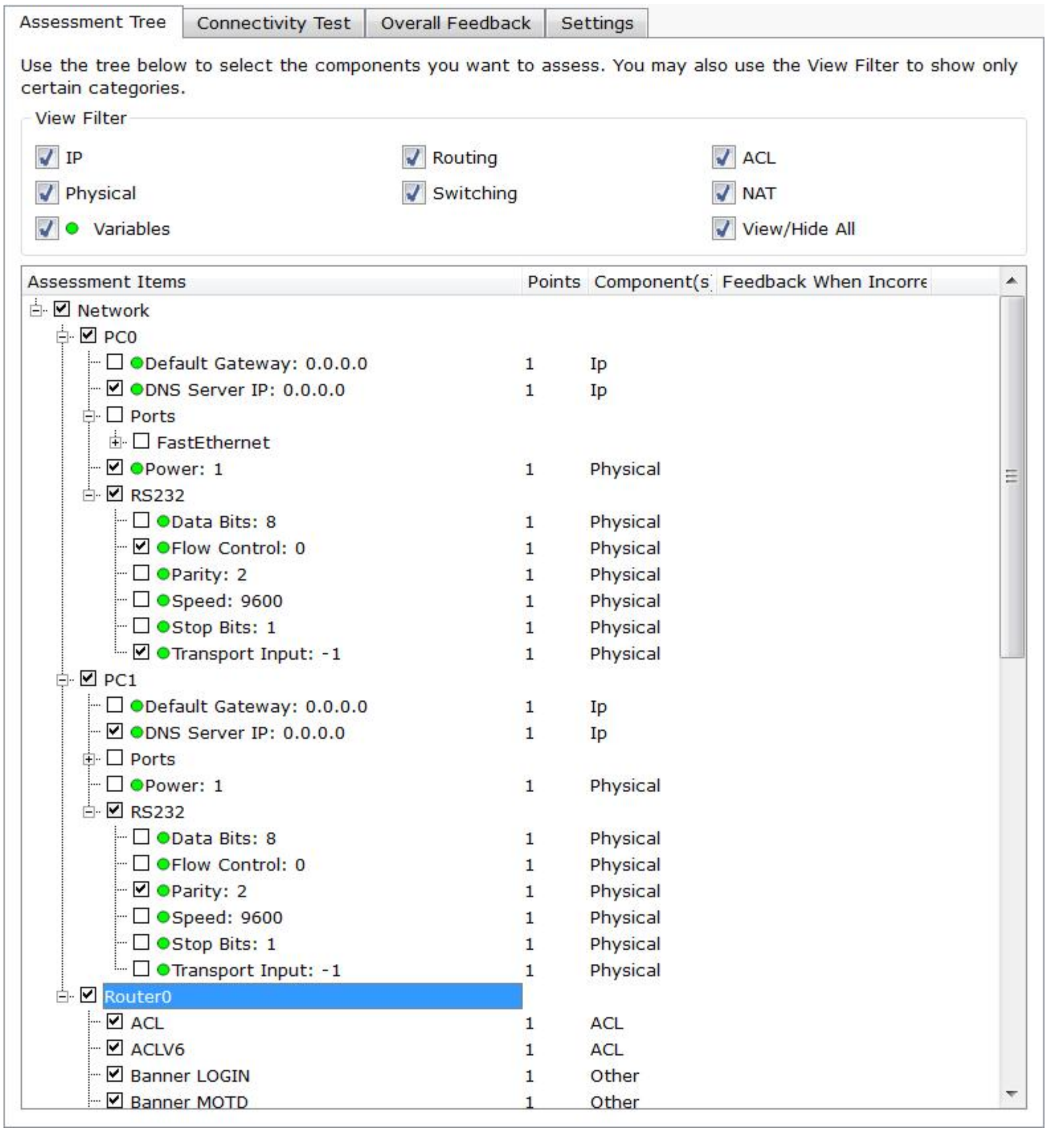


Figure 2.3 – Packet Tracer's Virtual Representation of the front of a PC

For the scope of this study, PT has two purposes. First, PT helps Cisco students learn how to create and troubleshoot computer networks that use Cisco products. Second, PT assesses learning through students' work from two activities using PT. According to Cisco Networking Academy, research is currently under way that is evaluating the validity of PT's assessment tree system. In particular, Cisco researchers are using evidence centered design (ECD) as a method for constructing assessment arguments (Behrens, Mislevy, Bauer, Williamson, & Levy, 2004; Mislevy, 2006; Mislevy, Steinberg, & Almond, 2003). ECD is a multi-layered approach to designing assessments by linking the desired student competencies to tasks that will provide evidence that the competencies were attained. ECD has been applied to Cisco training tools, such as *Networking Performance Skills System* (NetPASS), resulting in an evidentiary argument to assess performance and make inferences about student learning (Behrens, Mislevy, Bauer, Williamson, & Levy, 2004; Frezzo & Stanley, 2005). Parts of ECD were used in this study to develop an assessment argument that links the instruction presented in the tutorial to the tasks required for the PT activity.

Sims and Learning Theories

Education researchers have mostly adopted the concept that learning takes place both within the brain and outside the brain, such as situated cognition and distributed cognition. Sims have the potential to be authentic, high fidelity, situated, and distributed experiences. An effective sim is considered a constructivist tool that has a surface structure and deep structure (Gredler, 2004; Jonassen & Rohrer-Murphy, 1999). Surface structure is the task at hand, such as having to build a virtual computer network. Deep structure refers to the psychological mechanisms involved with carrying out the exercise. In sims, deep structure is complex because the “basis for a simulation is a dynamic set of relationships among several variables that reflect authentic causal or relational processes” (Gredler, p. 573). Thus, sims provide students opportunities to build a conceptual understanding of the system that is being represented. Mistakes, such as killing a patient or crashing a plane, have no consequences in sims; thus, learners can test their abilities and stretch their knowledge and skills. Similarly, sims allow students to see misconceptions in their understanding.

Sims provide "legitimate peripheral participation" in an authentic system, such as flying a plane or performing CPR (Lave & Wenger, 1991; Rieber, 2005). Therefore, sims are considered to be s*ituated* learning experiences (Dede et al., 2005; Dieterle & Clarke, 2007). For example, Chris Dede at Harvard developed *River City* using situated cognition theory. River City is designed so that students learn about science by acting like scientists in a virtual world. Students use their avatars to participate in an online community where other students and artificially intelligent agents use the scientific method to solve ill-structured, authentic problems. Authenticity is critical to an effective learning environment, especially sims (Herrington, 2007; Herrington & Oliver, 1998). In addition, Gredler (2004) stated that of critical importance to sims is that they are developed with high fidelity. In other words, a sim needs to be a realistic approximation of a complex reality.

Sims are also consistent with *distributed cognition* theories on human learning (Perkins, 1993; 1995). Distributed cognition theorists stated that cognitive processes, such as perception, learning, reasoning, and memory, occur outside and inside the head of an individual. Cognitive processes are distributed physically, socially, and symbolically between individuals and the tools they use. Symbolic distribution of cognition includes the sharing of symbol systems, such as mathematical equations or specialist language used only within an industry or job. PT is an example of symbolic distribution evident by the common symbols, icons, and language shared by Cisco Network Academy students and instructors.

Cognitive Load

For the purpose of this study, the primary theoretical framework used is cognitive load which is based on cognitive psychology and states that humans have limited capacities to process information (Sweller, 1988; Sweller & Chandler, 1994; Sweller, van Merrienboer, & Paas, 1998). This has important implications for the design of computer-based learning environments. In particular, cognitive load theory directly affects how pedagogically effective a sim is because overloading a student's cognitive resources has been found to decrease learning (Mayer, 2005; Sweller, 2005). First, a discussion of cognitive load's background is needed.

Cognitive Load Foundations

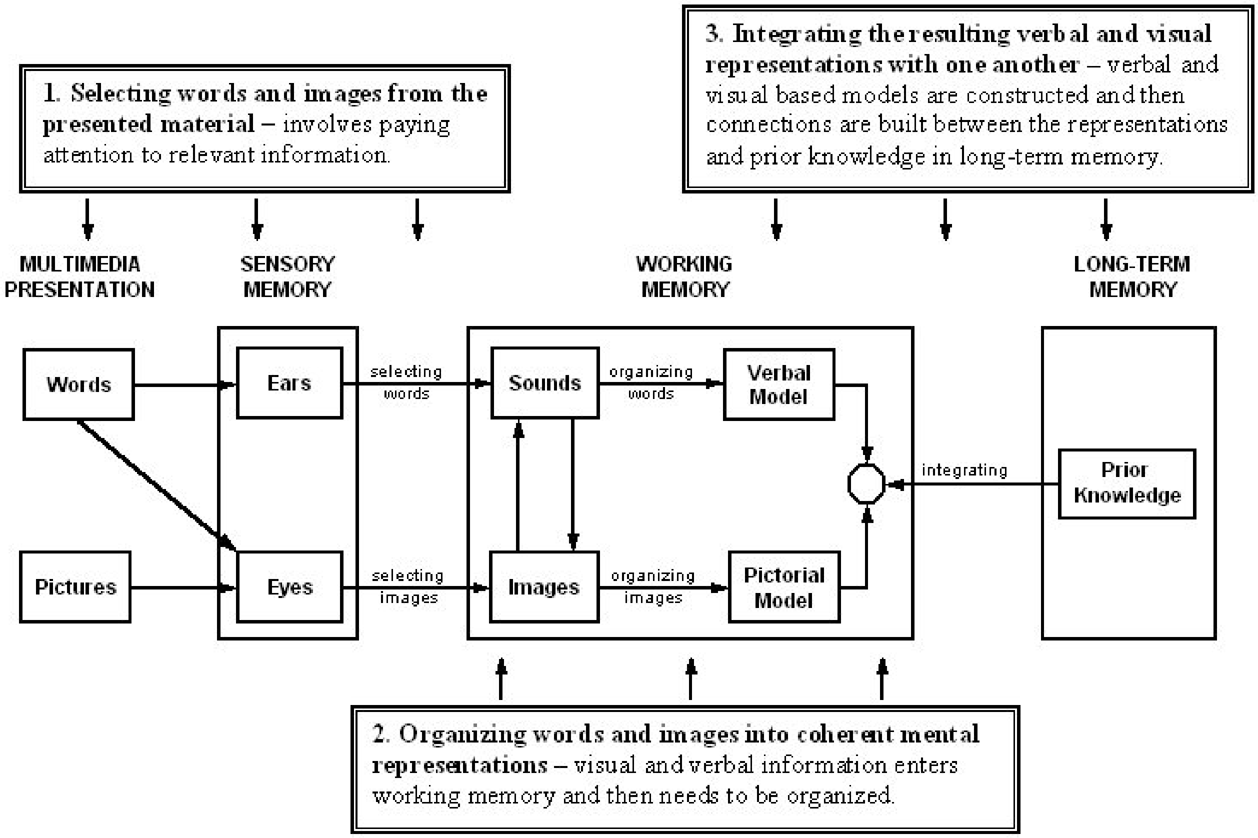
Baddeley (1986) postulated that working memory has a limited capacity and can only process a few pieces of information at one time. According to Sweller (1988), there are two major limitations of working memory when dealing with novel information: capacity and duration. This is consistent with Miller's (1956) finding that the human working memory capacity is limited to hold approximately seven elements of information. Similarly, Peterson and Peterson (1959) found that without rehearsal almost all contents of working memory are lost within about 20 seconds.

Paivio's (1986) *dual-code theory* stated that visual and auditory channels exist in which humans process information. These two channels represent verbal and pictorial sub-systems. Penny (1989) provided evidence that appropriate use of both sub-systems can increase working memory because capacity can be increased by using auditory and visual working memory together rather than using one or the other alone. The information being processed in each channel, however, needs to be related to be effectively processed or understood. This aspect of the theory could be tested in future research, although it is a logical assertion.

Information Processing Model

Both Baddeley (1986) and Paivio (1986) used an information processing model of learning. According to this model, humans first perceive information from the outside world which then enters the sensory system through the visual and auditory channels where the input resides temporarily in sensory memory. When humans register the information, it enters the working memory stage for active consciousness. Sounds are registered and organized and then become a verbal model. Images are registered and organized and then become a pictorial model. Meanwhile, humans also access prior knowledge from their long-term memory and integrate it with the information in the verbal model and pictorial model. Meaningful learning occurs when humans are at this point of integration, see Figure 2.4.

Figure 2.4 – Information Processing Model (Mayer, 2001)



This information processing model with integration at the center is consistent with the *Memory-Consolidation Hypothesis* which states that information or a stimulus enters a temporary memory store and must be integrated into long-term memory (Mayer, 2001). Information in the temporary memory store can be lost due to distractions, such as extraneous information on-screen. Neurological research has shown support for this model and claim that multi-modal input (combinations of visual and auditory stimuli) can lead to increased synaptic firing followed by re-wiring and increases in synaptic strength (Lisle, 2007; Whelan, 2007). The result of the process creates long-term potentiation or long-term memory, i.e., learning.

Three Types of Cognitive Load

Sweller and colleagues have developed a triarchic cognitive load theory in which cognitive load is comprised of three types of cognitive load: Intrinsic, Germane, and Extraneous (Sweller & Chandler, 1994; Sweller et al., 1998). In addition, neurological evidence is beginning to emerge that supports the triarchic cognitive load theory (Whelan, 2007). The three types of cognitive load are considered additive in that they are each independent but are all inter-related and together comprise a general cognitive load level.

Intrinsic cognitive load is the inherent cognitive resource caused by the complexity of learning content (Sweller, 1988). There is an inherent difficulty in learning that is associated with how complex information is for a learner. For example, reading a law school textbook is going to be more complex than reading a 5^{th} grade history book. The higher the complexity relative to the learner's knowledge level, the higher the intrinsic cognitive load.

Germane cognitive load is the use of relevant cognitive resources caused by the learner's investment on schema construction and automation. "Effective instructional methods encourage learners to invest free processing resources to schema construction and automation, evoking germane cognitive load" (van Merriënboer & Sweller, 2005, p. 152). Germane load requires meaningful mental effort, such as learners making sense of a concept, solving a problem, or practicing skills.

Extraneous cognitive load is the irrelevant cognitive load caused by the medium, layout or structure of instruction. Extraneous load takes the learner's attention away from the task at hand and, more importantly, away from the process of schema construction (van Merriënboer & Sweller, 2005). Examples of extraneous load are apparent in poorly designed web sites that overload the page with so much information that the user quickly abandons the site by clicking “back” on the browser. In complex sims, extraneous cognitive load can take the form of too many buttons, sliders, features, or displays on the sim’s user interface. For novices, a complex interface may produce extraneous overload; however, for an expert, the sim’s interface may be too simplistic. Expertise is commonly found as a moderator of extraneous cognitive load (Kalyuga et al., 2003).

Experts vs. Novices

Cognitive load theory applies to novel information or stimuli. “Neither the duration nor capacity limitations attached to novel information received from sensory memory applies to information from long-term memory” (Sweller, 2006, p. 24). In other words, experts in a given subject or skill are much less likely to experience cognitive overload in their area of expertise. Their well developed schemas in long-term memory allow for processing more information than a novice who must first build such schemas of the domain.

Experts differ from novices not only by the former’s schema but also because experts are more effective at combining simple ideas into complex systems (van Merriënboer & Sweller, 2005). Thus, do developers need to design sims differently for novice vs. experts? Some researchers say “yes.” For example, the expertise reversal effect states that “instructional methods that work well for novice learners may have neutral or even negative effects when expertise increases” (Kalyuga et al., 2003; van Merriënboer & Sweller, 2005, p. 149). “This effect necessitates the formulation of instructional strategies that make the application of particular instructional methods dependent on learners’ expertise” (van Merriënboer & Sweller, 2005, p. 152). In other words, sims need to be developed for varying levels of expertise; although, developing multiple versions of the same sim is impractical. A more likely solution is to optimize sim interfaces for different levels of prior knowledge.

Lee, Plass, and Homer (2006) examined how cognitive load in visual displays of sims should be optimized. They found that intrinsic and extraneous cognitive load in sim interfaces can be manipulated and that learners' prior knowledge moderates the effectiveness of these load manipulations. Their sample consisted of 257 middle school students using a sim modeling the ideal gas law. Visual complexity was manipulated by having a low complexity (2 screens) and a high complexity (1 screen) group. The low complexity group promoted comprehension and transfer, especially for low prior-knowledge learners. An expertise reversal effect was found for learners with high prior general science knowledge.

The expertise reversal effect occurs when instructional methods, such as a complex sim, that are highly effective for students with low levels of knowledge can actually impair or even reverse learning for students with high levels of knowledge or expertise (Kalyuga et al., 2003). Kalyuga and colleagues reviewed an extensive amount of empirical literature examining the interaction between instructional techniques and the level of a learner’s expertise. They found evidence of the expertise reversal effect replicated in multiple settings, and they concluded that their findings confirmed that low prior knowledge students need to be scaffolded by using sims that have little extraneous or non-task related information on-screen. Given these findings, further research needs to examine the relationship between a sim's cognitive load and the learner's prior domain knowledge.

Measurement of Cognitive Load

This section reviews strengths and weaknesses of various methodologies used to measure cognitive load. Paas and colleagues have used cognitive load measures including expert opinion, task analysis, self-report scales, performance-related data, and psycho-physiological data such as heart rate, pupil dilation, and galvanic skin response (Paas, Tuovinen, Tabbers, & Van Gerven, 2003); however, according to Whelan (2007) these measures do not show the distinctions between the different types of cognitive load.

Researchers have examined physiological measures, including electrocardiogram (ECG) and heart rate, relationship to cognitive load but found no conclusive evidence (Paas et al., 2003). In more recent research, Whelan (2007) found evidence of the three types of cognitive load. "Cognitive load theory has a basis in functional neuroanatomy, and functional Magnetic Resonance Imaging (fMRI) techniques will allow us to accurately observe the properties of certain brain functions related to different types of cognitive load" (p. 5).

Measurement of Cognitive Overload in Sims

Cognitive load is an effective framework for examining learning in computer based environments because of the potential for experimental control, factor manipulation, and learner assessment; however, measuring cognitive load remains an issue due to its dynamic and internal nature. There have been numerous methodologies used for measuring cognitive load in sims and multimedia learning. These techniques have included: (a) subjective self reports, (b) a secondary-task technique, and (c) a psychophysiological measure.

Brunken and colleagues used a dual task approach to measuring cognitive load. A learner must respond to a secondary task while attending to the primary task thereby inducing memory load (Brunken, Plass, & Leutner, 2003). The dual task approach can give immediate real-time indications of cognitive load, and the use of a within subject design makes the measurement of cognitive load independent of individual differences which would affect a between subject design (Whelan, 2007). However, there are weaknesses to the dual task approach, including learners using new strategies to work with the secondary task and thereby influencing performance (Meshkati & Loewenthal, 1988).

Subjective measures are frequently used to measure cognitive load. These measures are frequently questionnaires comprising one or more scales, such as six or nine point Likert scale, in which learners can indicate their level of mental effort, fatigue or frustration experienced. Researchers have recently been using a *learning efficiency score* which attempts to quantify the relationship between cognitive load and performance (Kalyuga & Sweller, 2005; Paas & van Merrienboer, 1993). High learning efficiency occurs when learner performance is higher than learner mental effort. This assumes that high efficiency is indicated by higher performance and lower mental effort (Paas et al., 2003). However, there are problems with learning efficiency scores. Such scores are counter intuitive because high mental effort is necessary for deep learning (Bloom, 1956).

Measuring the Intrinsic, Germane, and Extraneous Cognitive Load

It has recently become standard practice for cognitive load researchers to measure all three types of load. Recent research by DeLeeuw and Mayer (2008) systematically measured the three types of load at eight time points during the session. Self-report scales were used to measure mental effort and a difficulty rating scale was completed at the end of the lesson. Response time to a secondary visual monitoring task was used as a behavioral measure of extraneous processing. The researchers manipulated the three types of cognitive load by adding redundant text (extraneous load), increasing the complexity of sentences (intrinsic load), and comparing transfer of knowledge (germane load). Learners who have high transfer scores are likely to have utilized germane load while integrating knowledge and constructing schemas. Additionally, they found low correlations between the measures of each type of cognitive load, which they consider evidence that the measurements are measuring different constructs.

DeLeeuw and Mayer (2008) recommended that “When the goal is to assess the level of extraneous cognitive load, RT to a secondary task appears to be most appropriate; when the goal is to assess the level of intrinsic cognitive load, mental effort ratings during learning may be most appropriate; and when the goal is to detect the learner’s level of germane cognitive load, a simple difficulty rating immediately after learning may prove most useful” (p. 234). This study employs all three forms of measurement stated by DeLeeuw and Mayer; however, RT is replaced with the log data from PT.

Measuring Intrinsic Load

Van Merriënboer and Sweller (2005) measured intrinsic cognitive load by the difficulty level of the domain since intrinsic load is influenced by the complexity of materials being learned. Kalyuga and Sweller (2005) used the following item to measure intrinsic load: "Please indicate how difficult the instruction/test you just took was by clicking on the appropriate degree of difficulty." Similarly, DeLeeuw and Mayer (2008) used a Mental Effort rating as a measure of intrinsic cognitive load: "please rate your level of mental effort on this part of the lesson." The item included a Likert-type response scale ranging from 1 (Extremely low mental effort) to 9 (Extremely high mental effort). This item was administered eight times throughout the duration of the study's lesson. Cronbach's alpha indicated a good estimate of reliability for the Mental Effort measure ($\alpha = .90$).

Measuring Germane Load

Measuring germane cognitive load should examine the level of the learners' schema construction and automation. Researchers have had difficulty measuring germane cognitive load; however, DeLeeuw and Mayer (2008) claim to have reliably measured germane cognitive load using one item for a Difficulty rating: "Please indicate how difficult this lesson was by checking the appropriate answer." A Likert-type response scale was used with responses ranging from 1(Extremely Easy) to 9 (Extremely Difficult).

Measuring Extraneous Load

Extraneous cognitive load is considered to decrease learning because the student is distracted or overloaded by stimuli and information. There are numerous methods for measuring extraneous cognitive load, including behavioral measures and self-report measures. Behavioral measures have included navigation errors and orientation problems because making errors while navigating a system takes away from germane and intrinsic load processing (Astleitner & Leitner 1996; Brunken et al., 2003, p. 56).

Summary

There is significant need to continue researching how to measure the three types of cognitive loads. By measuring each, we can examine relationships between each: will germane load decrease with high intrinsic load or high extraneous load, or will germane load increase with low extraneous load? Further, by measuring each type of load we can examine their relationships to performance. However, measuring each type of cognitive load is difficult (Whelan, 2007). For example, subjective self-reports are situated in a setting and context that affects the rater's score as much as the actual task difficulty. Individual differences are a factor because students may have various interpretations of the meaning of the self-report questions. Additionally, there is a lack of consistency between performance ratings and subjective ratings of workload, difficulty, and effort. DeLeeuw and Mayer's (2008) self-report measures for Mental Effort and Difficulty only included one item; however, the item was administered eight times during the session. In this study, Mental Effort and Difficulty items were administered after the PT activity.

Cognitive Theory of Multimedia Learning

Cognitive load theory is the foundation of Mayer and colleagues research on learning in multimedia environments. The cognitive theory of multimedia learning (Mayer, 2001) uses the information processing model of learning discussed above. Based on this cognitive architecture, it has repeatedly been found that students learn better when words are presented as spoken text with pictures or animations compared to being presented with printed text with pictures or animations (Mayer, 2005). The rationale being that when pictures and words are both presented visually, the visual channel is overloaded (Sweller & Chandler, 1994). When words are processed through the auditory channel, the visual channel is available to process the pictures or animation (Mayer & Johnson, 2008). However, the previous research findings show evidence of a modality effect mostly with materials that show animations the physical process of how lightning forms and how pump brakes work. Appendix A provides an overview of the cognitive theory of multimedia learning's principles and effect.

Modality effect

The modality effect essentially states that people learn deeper from animation and narration than from animation and on-screen text (Mayer, 2001). This effect has been tested and supported more than two dozen times (Mayer, 2005; Sweller, 2005). Ginns' (2005) meta-analysis of the modality effect examined 43 independent effects (39 between-subjects designs, 4 within-subjects designs). Modality effect was overwhelmingly supported thus indicating that there are instructional benefits to presenting information across modalities. "Across a broad range of instructional materials, age groups, and outcomes, students who learned from instructional materials using graphics with spoken text out-performed those who learned from a graphics with printed text" (Ginns, 2005, p. 326).

The modality effect has been found to be strong for measures of transfer but not for retention (Ginns, 2005; Mayer, 2001). Printed text and an animation requires the learner to split their attention leading to extraneous processing in the visual channel. This reduces the cognitive resources available for intrinsic and germane processing. Intrinsic processing is used first when selecting and attending to information which often leaves little resources left for germane processing (Sweller, 1988). Thus, a learner who has to read and watch a lesson may be utilizing cognitive resources for intrinsic processing rather than germane processing and schema construction. However, when the learner is able receive a voice narration and the animation the visual channel is not split reducing extraneous load and leaving more resources available for germane processing. Accordingly, learners who are presented with voice narration and animation have been found to score better on measures of transfer compared to learners who received text and animation (Harskamp, Mayer, & Suhre, 2007).

Ginns also discussed two moderators that affected the modality effect: (1) level of element interactivity and pacing of presentation, and (2) between certain fields of study. Ginns (2005) found that the modality effect was larger in studies where pace of presentation was set by the system compared to the self-paced presentations. In this study, the tutorials are not interactive; however, PT is very interactive. Additionally, the tutorials are not self-paced because of the need for a controlled experimental study. As Ginns found, pacing can moderate the modality effect.

Redundancy principle

Research has shown that students learn better from multimedia lessons containing graphics and narration compared to graphics, narration, and redundant on-screen text (Mayer, 2001, 2005). Sweller (2005) states that redundancy effect is when “the elimination of information from instructional material results in improved learning” (p. 161). Two conditions are usually examined in redundancy effect experiments with one condition consisting of a full set of instructional material and a reduced set of material. If learners in the reduced set perform better on learning measures compared to the full set then the redundancy effect has been obtained.

Findings indicate that there is a delicate balance between how much on-screen text is too much. Mayer and Johnson (2008) found that short on-screen labels guided the cognitive process of selecting relevant words and images while not creating extraneous processing. However, in numerous other studies, Mayer found that having a complete text of the narration shown as on-screen text resulted in decreases in both retention and transfer test performance. Further, their study used static images, such as in a PowerPoint slideshow, rather than animation. Thus, their findings cannot be generalized to animations such as the tutorial tested in this experiment.

Assessment

Assessment is tough with sims because they are complex environments. Clarke and Dede (2007) describe three types of data for conducting assessment in such environments, e.g., virtual worlds, sims, games, etc. First, *contextual data* includes demographic information about the learner. Second, *assessment data* are considered measurements of desired knowledge, skills, and abilities. Retention and transfer tests are examples of assessment data that are commonly used in modality research, especially for testing the redundancy effect (Ginns, 2005). A retention test used frequently in Mayer's research was an open-ended question stating "Please write down an explanation of how lightning works" (Mayer, 2001; Mayer &Johnson, 2008; Moreno & Mayer, 2002). The retention tested is intended to measure the student's ability to recall information from the animated lesson on how lightning forms in the sky. The transfer test used in the lightning line of studies included four open-ended questions (p. 382):

- "What could you do to decrease the intensity of lightning?"
- "Suppose you see clouds in the sky but no lightning. Why not?"
- "What does air temperature have to do with lightning?"
- "What causes lightning?"

In Mayer's research, both the retention and transfer tests were scored by assigning 1 point for listing an acceptable answer. A list of acceptable answers was developed for each question. There were 16 acceptable answers on the retention test and 12 for each item on the transfer test. Retention tests have been considered tests of lower level cognitive processing of the information; whereas, transfer tests are believed to measure higher levels of processing, such as application, evaluation, and synthesis (Bloom, 1956; Ginns, 2005; Mayer, 2001, 2005).

Third, *active data* are students' actions and behaviors as they learn via "mediated interaction." For example, active data include team chat transcriptions, work products from activities, notes, and log data of a user's movements and click-path while using a software or web site. Another type of data relevant to this study are self-report measures, such as a usability survey.

Rieber (2005) stated that when evaluating learning in a sim, a researcher needs to examine *explicit* and *implicit* learning. Explicit learning is measured with traditional multiple-choice tests, which is essentially assessment data (Clarke & Dede, 2007). Implicit learning uses more behavioral measure, such as the ability to complete the activity like fixing PT's broken network. Rieber's implicit learning is parallel to Clarke and Dede’s active data. In this study, both explicit and implicit learning were measured using assessment tests (retention) and active data (PT activity transfer test).

Summary

This review of the literature on sims, cognitive load theory, and the cognitive theory of multimedia learning revealed opportunities for research. First, evidence continues to build supporting the modality effect (Ginns, 2005; Mayer, 2005); however, there is no research on the modality effect in the context of a tutorial that teaches novice students how to use a complex simulator. How should instructional content in a sim tutorial be delivered in terms of text, voice narration, or both? Second, novices need non-expert sims; yet, building separate novice and expert sims is unrealistic. Thus, what effect does simply restricting access to certain parts of a sim’s interface have on performance of a task? There is no research addressing this question, yet it provides a simple solution to a common problem. Finally, is there an interaction between a tutorial’s modality and restriction of a sim’s interface which affects the user performance?

Chapter 3 Methods

The purpose of this study was to examine how manipulating the modality (text-only, voice-only, voice+text) of a tutorial and how restricting (restricted vs. unrestricted) part of a sim's interface would affect learning on retention and transfer tests. This study addressed relevant instructional technology design questions, such as how to design tutorials for complex sims and what effect restricting a sim's interface has on learning for novice students. In short, I used PT and a PT tutorial to explore the cognitive theory of multimedia learning and cognitive load theory.

Participants and Design

Participants

The total sample included 81 subjects from the University of Texas at Austin College of Education subject pool: 25 males (31%) and 56 females (69%). In terms of grade classification, the study was composed of 45 seniors (56%), 19 juniors (23%), 11 sophomores (13%), three freshmen (4%), and three graduate students (4%).

Experiment Design

Subjects were randomly assigned to six conditions that were already setup at each computer: text-only restricted (TR), voice-only restricted (VR), voice+text restricted (VTR), text-only unrestricted (TU), voice-only unrestricted (VU), and voice+text unrestricted (VTU). A 2x3 between-subjects design (See Table 3.1) was used to test for differences among the groups.

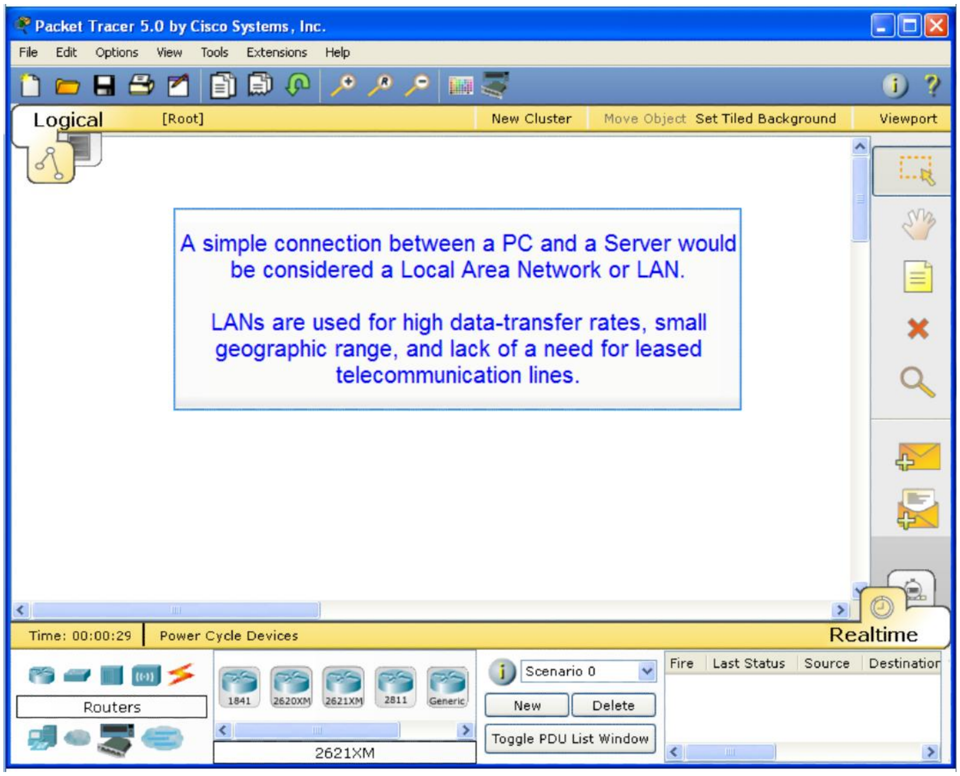
Table 3.1 – Experiment Design: Tutorial Modality vs. Interface Restriction

	Text-only	Voice-only	Voice+Text
Restricted	TR	VR	VTR
Unrestricted	TU	VU	VTU

Materials and Apparatus

The computer-based materials consisted of an online survey for a pretest and a post-test of retention, a 10-minute tutorial, and PT. The online surveys were administered using Survey Monkey. The tutorial was developed using Adobe Captivate. The final version for each modality condition was exported to a Flash file and embedded in a Firefox browser window. The tutorial consisted of content from Cisco Networking Academy's *My First PT Labs*. This content covered how to create, configure, and test a network. I created a three minute introduction to computer networking to provide novices with a basic understanding of computer networking. For experimental control, the tutorials were made without Playback Control bars, which allow a user to stop, pause, rewind, and fast-forward through a tutorial. Figure 3.1 below shows the text-only tutorial.

Figure 3.1 – Screenshot of the Text-only Tutorial



PT Activity

For experimental control, students were given 10 minutes to accomplish the PT activity. PT's interface was restricted for half of the subjects and unrestricted for the other half. Figure 3.2 shows the restricted or "locked" parts of PT's interface in red. Instructions are presented to the student upon opening PT, see Figure 3.3 below. Subjects were instructed to pay attention to the timer and score in the instructions box, shown in red.

Figure 3.2 – PT Restricted: locked parts of interface shown in red

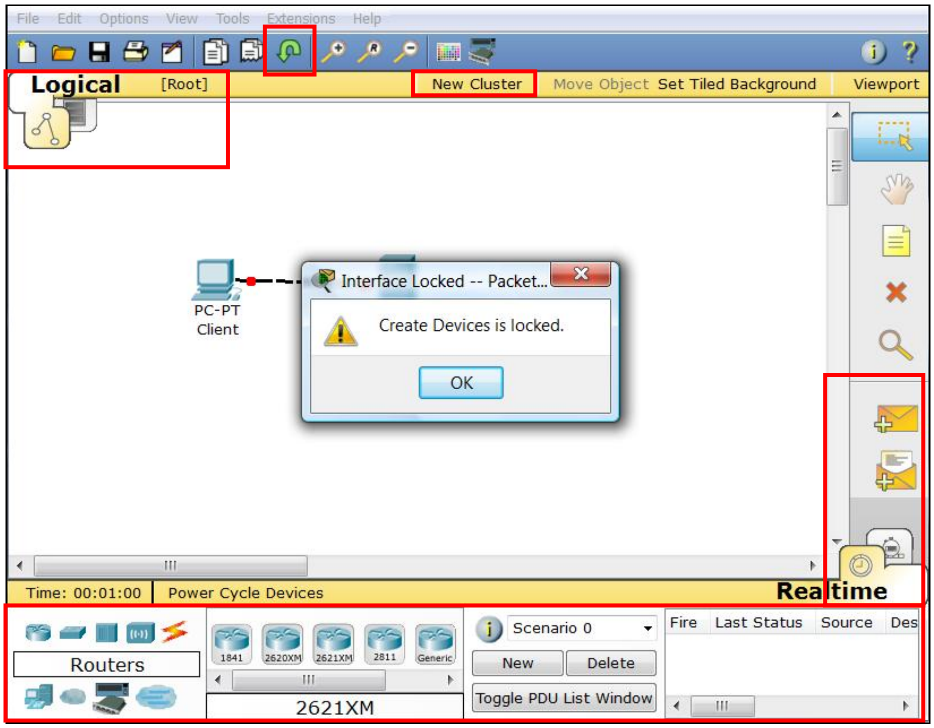
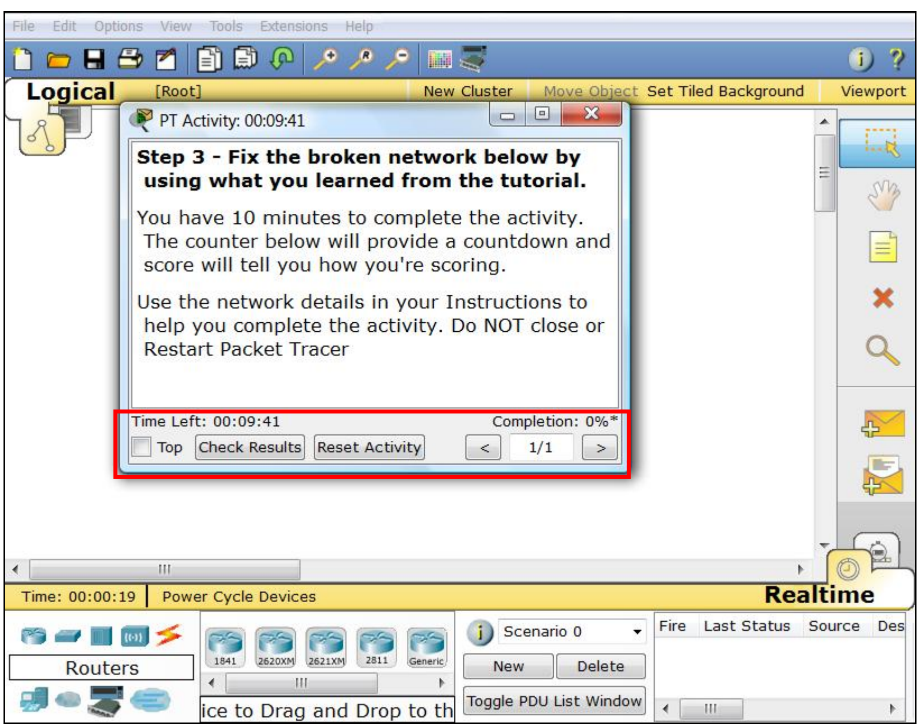


Figure 3.3 – PT's instructions delivered upon at start. Timer and score in red.



Measures

Numerous types of data were collected in this study: a pretest of computer networking knowledge, a retention test of information from the tutorial, a transfer test applying information from the tutorial, and self-report items measuring affective responses to the tutorial and sim as well as items that measure the three types of cognitive load. The retention test was a declarative test over the first half of the tutorial, which covers basic computer networking terms and concepts. The PT activity was a transfer test of the student's ability to apply the process of setting up a network, which was shown in the second half of the tutorial. This process was demonstrated through an animation that accompanied the text and/or voice narration depending on the modality condition. The animation is more important in the application of information because it illustrates a process, e.g., fixing the network. Table 3.2 lists each measure, when the measure was administered, and the measure's data type.

Table 3.2 – Assessments, Measures, & Questionnaires

Measure	Description	Pre	Post	Data Type
Pretest	Computer networking knowledge. 10 MC items. (Appendix B)	X		Prior Knowledge
Post-test	PT Trouble Shooting activity		X	Transfer test
	Retention from tutorial. 14 MC items. (Appendix C)		X	Retention
Cognitive Load	Interface restriction 3 items using a 5-point Likert scale. (Appendix D)		X	Extraneous Cognitive Load
	Mental effort rating. 3 items using a 5-point Likert scale. (Appendix E)		X	Intrinsic Cognitive Load
	Difficulty rating. 3 items using a 5-point Likert scale. (Appendix F)		X	Germane Cognitive Load

Pretest

The pretest consisted of 10 multiple-choice items that assessed the subject's prior knowledge of computer networking (see Appendix B). Additional questions included gender and grade level. Consistent with previous cognitive theory of multimedia learning research (Mayer, 2001), tests of retention and transfer were collected in this study.

Retention post-test

The retention test included 14 multiple-choice items and assessed students comprehension and recall from the tutorial. The items focused on information presented at the beginning of the tutorial, especially vocabulary and basic concepts (see Appendix C).

PT Activity transfer test

The PT activity required that subjects transfer what they learned from the tutorial to fixing the broken network shown on PT's workspace. Subjects used PT for 10 minutes with the task of fixing a network that needed to be properly configured. Subjects started with a server icon and a computer icon on the PT workspace. Fixing the network consisted of nine steps. For the PC, subjects had to (1) turn the power on, (2) enter the IP address, (3) enter the DNS server, and (4) click Port Status. For the server, subjects had to (5) turn the power on, (6) enter the IP address, (7) click DNS to on, (8) click Port Status, and (9) click HTTP to on. The rationale used for this activity was that if the tutorials were in successful teaching students how to fix a network using PT, then they successfully transferred knowledge from the tutorial to the PT activity. Parts of the ECD assessment framework (Mislevy, 2006) were used in this study to develop an assessment argument that links the instruction presented in the tutorial to the tasks required for the PT activity.

Appendix G provides an ECD worksheet that details the task-analysis from the tutorial content.

Reliability of measures

The internal consistencies of the measures were low given the small number of items on the pretest and the post-test. The pretest had 10 items with a Cronbach's alpha of .37, and the retention test had 14 items with a Cronbach's alpha of .49. The low alphas for these two measures created potential for inaccurate results due to weak measurement. Thus, caution was used when interpreting the results that used these measures. This is discussed more in the Results, Discussion, and Limitations sections. The the PT activity transfer test score was a composite score determined by the number of tasks completed correctly.

Self-Report measures

Self-report items were used for a treatment integrity check and to assess subjects' attitudes towards PT and the tutorial. The three types of cognitive load were measured using Likert type items. The measures were adapted from DeLeeuw and Mayer's (2008) recent measures of the three components of cognitive load. Subjects' affective responses to PT and the study overall were also measured using Likert type items. The self-report items were administered within the same online survey as the retention test.

Extraneous cognitive load items

Extraneous cognitive load was measured using self-report items that asked subjects about various aspects of PT, including usability questions. These questions used a five-point Likert scale ranging from Strongly Disagree (1) to Strongly Agree (5). Items that all subjects received included: "Packet Tracer was fun," "Packet Tracer's interface (visual display) was easy to use," "Packet Tracer was frustrating," and "Packet Tracer's interface was overwhelming." Items that only the Restricted groups received included: "The 'Locked' messages distracted me" and "Not being able to access a restricted part of Packet Tracer made me curious about why I was being restricted." Appendix D lists the extraneous cognitive load questions which could also be considered as usability questions.

Intrinsic cognitive load items

Intrinsic cognitive load was measured with four self-report items that asked subjects to rate their mental effort during parts of the study. Two questions asked subjects to rate their mental effort ranging from extremely low to extremely high during the tutorial and while using PT. The next two questions asked subjects to rate the complexity of the tutorial and PT ranging from Extremely simple (1) to Extremely complex (5). Appendix E lists the four items intended to measure intrinsic cognitive load. DeLeeuw and Mayer's (2008) intrinsic cognitive load measures had subjects rate their mental effort eight times during a session. This study did not repeatedly have subjects rate their mental effort due to time constraints and administrative complexities. However, other questions were developed to address mental effort.

Germane cognitive load items

Germane cognitive load was measured with three self-report items that asked subjects to rate the difficulty of parts of the study on a scale ranging from Extremely easy (1) to Extremely difficult (5). The items included how difficult it was to understand the tutorial, remember information from the tutorial, and using PT. DeLeeuw and Mayer (2008) had subjects rate the difficulty of parts of the activity as a way to measure germane cognitive load; however, while DeLeeuw and Mayer used one item, this study used three to increase the reliability of the measurement.

Affective items

Subjects rated various aspects of their experience using PT and during the study. The three self-report items all used five-point Likert scales; however, each item had a different response scale. The first item asked “If you had a chance to continue exploring Packet Tracer, how eager would you be to do so?” The response range included Very Uneager (1) to Very Eager (5). The second item asked “How interesting is this material?” with responses ranging from Very Uninteresting (1) to Very Interesting (5). The third item asked “How much effort did you put into this study?” with responses ranging from No Effort (1) to All My Effort (5).

Apparatus

All experimental sessions were conducted in a computer lab at the Learning Technology Center (LTC) within the George I. Sánchez building at the University of Texas at Austin. The computer labs have approximately 25 PC computers with a 1.8 GHz Intel processor, 1 GB of RAM, a 17-in. LCD monitor, and a 150 GB hard drive. For the voice-only groups, USB headphones were provided by the LTC.

Procedure

Pilot testing & data collection

I pilot tested the apparatus and measures prior to running the study to correct technical problems and improve the measures. First, a group of four graduate students pilot tested the study and provided feedback on how to simplify the study’s process. Two sessions with 27 subjects total were dedicated to pilot testing with improvements made to how subjects selected which tutorial condition they watched and which PT activity condition they opened. In addition, modifications were made to the pretest and post-test by examining which distracters were not working and which items were poorly worded. Formal data collection began with Session 3 on March 9th, 2009 (see Table 3.3).

Table 3.3 –Data collection sessions, dates, and number of subjects in each session.

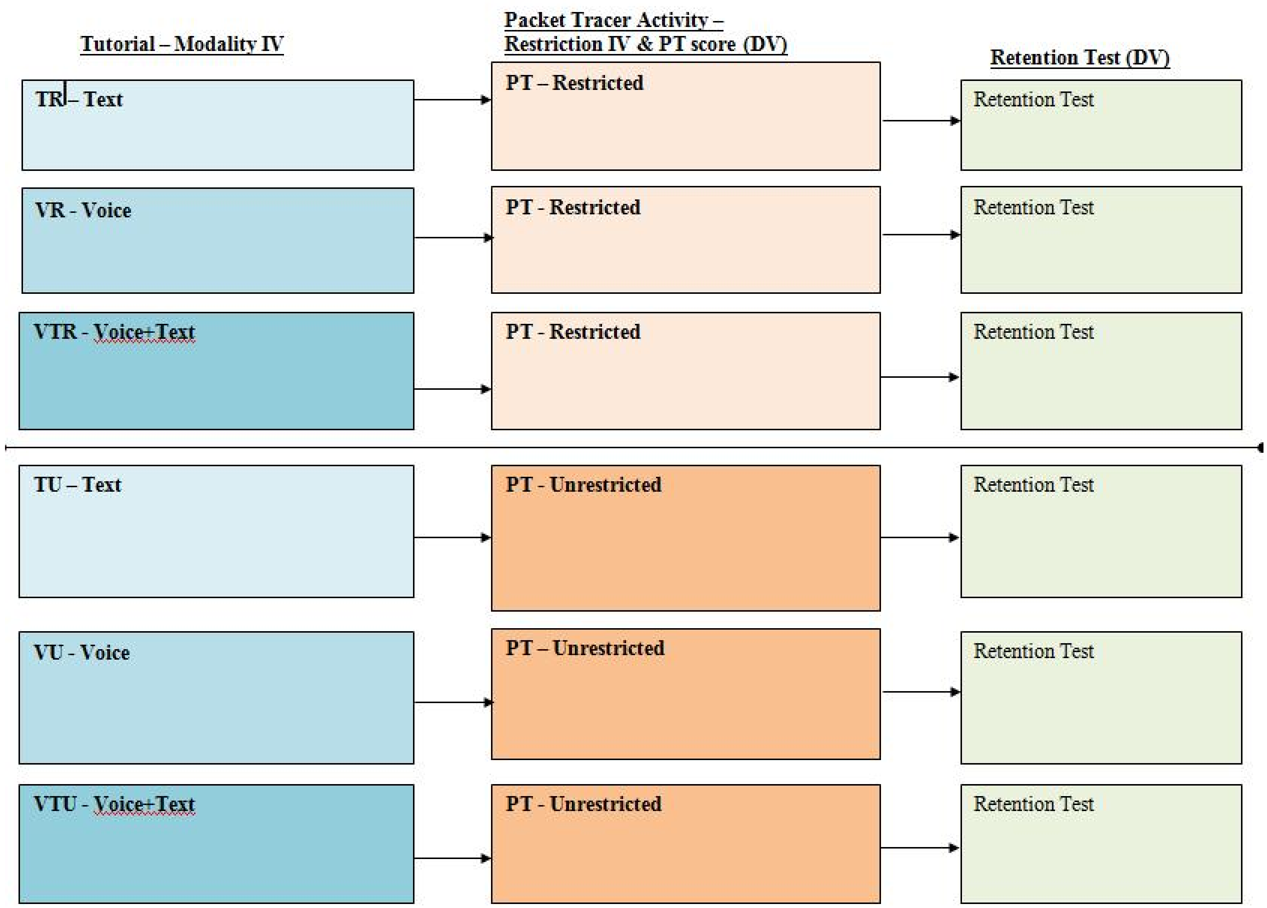
Session number	**Date**	**Subjects in each session**	**Description**
Session 1	March 6, 2008	14	Pilot test
Session 2	March 6, 2008	13	Pilot test
Session 3	March 9, 2008	13	Formal data collection begins
Session 4	March 9, 2008	14	
Session 5	March 9, 2008	14	
Session 6	March 10, 2008	4	
Session 7	March 10, 2008	9	
Session 8	March 10, 2008	11	Discovered PT scoring issue. Lost data for Sessions 3-8 due to unreliable PT scores.
Session 9	March 11, 2008	6	Started telling students to not change PT computer names.
Session 10	March 11, 2008	11	
Session 11	March 11, 2008	9	
Session 12	March 11, 2008	16	
Session 13	March 11, 2008	15	
Session 14	March 12, 2008	7	
Session 15	March 25, 2008	13	
Session 16	March 25, 2008	4	

Study session procedure

Participants were randomly assigned to computers in the computer lab. Prior to the subjects arriving, each computer was setup which included logging in, instructions (Appendix H) with a condition identification code, and a consent form (Appendix I). The instructions provided a list of four steps the subject must accomplish: pretest, tutorial, PT activity, and post-test. The instructions also provided information necessary to setup the network, such as an IP address. Depending upon their random assignment to one of the three modality conditions, subjects then watched and listened to the 10 minute tutorial.

Next, they used PT for 10 minutes to fix a broken network by configuring a virtual PC and server with IP addresses, DNS settings, and by clicking on buttons demonstrated in the tutorial. Following completion of the PT activity, subjects returned to Survey Monkey and completed the post-test. Subjects checked their names on a roster and at the end of the session I gave them a certificate of completion (Appendix J). Subjects never provided any personal information. Figure 3.4 below illustrates the sequence of events and the conditions of the study: tutorial, PT activity, and retention test followed by self-report items.

Figure 3.4 – Study Session Procedure



Research questions

RQ1. Given research on the modality effect, how does the "My 1st PT Lab" tutorial instructional delivery modality (text-only, voice-only, and voice+text) affect performance on dependent measures?

RQ2. What effect does PT's interface being restricted have on dependent measures?

RQ3. Is there an interaction between modality and interface restriction?

Hypotheses

Based on existing literature, it was hypothesized that the high load condition would overload the visual channel, thus resulting in poorer performance compared to the control and low conditions (Mayer, 2001; Sweller, 1988). In addition, it was hypothesized that high prior knowledge students will show higher performance gains from pretest to post-test compared to students with low prior knowledge (Mayer, 2005).

H1. Research states that voice-only would outperform the other two conditions (Mayer & Johnson, 2008; Moreno & Mayer, 2002).

H2. There is no available literature comparing the effects of a sim's interface being restricted or unrestricted. However, it was expected that because these PT users are extreme novices, the restricted group will experience less cognitive overload and therefore outperform the unrestricted group (Kalyuga et al., 2003).

H3. It was predicted that there would not be a significant interaction between modality and interface restriction. However, there may be an additive effect. It was also predicted that the best condition would be voice-only, restricted.

Chapter 4 Results

Loss of Data due to issues with PT Activity

There were procedural issues with the PT activity throughout the entire data collection process that reduced the number of subjects whose data could be used for analysis. First, a major scoring issue became apparent midway through the experimental sessions. The issue was identified after the 8th Session or 65th subject. For the PT activity, subjects were tasked with fixing a broken computer network between a PC and server. Fixing the network required configuring both the PC and the server, which entailed inputting IP addresses, DNS server settings, and making sure that certain buttons were clicked to "On." Changing the configuration name of the PC or server was not part of fixing the network for the PT activity; however, the Cisco tutorial used in the study demonstrated how to change these names. As a result, many subjects did change the configuration name on the PC or server.

The issue is that if subjects changed this configuration name for either the PC or server, they received no points for that item even if everything else was configured properly. In other words, if a subject changed the name on the PC from "Client" to "PC1" then that subject would lose all points for the PT. Rather than getting all four possible points for the PC, the student would receive no points even if the subject had correctly configured the system. Similarly, if the name of the server was changed from "Web Server" to something else, then all five points were lost for that part of the PT activity.

The PT activity was developed in collaboration with individuals from Cisco Networking Systems; however, this important scoring detail was not discussed and was therefore not discovered until the end of Session 8 when I investigated why the PT activity scores were consistently so low. Sessions 1 and 2 were used for pilot testing, and numerous improvements to the study were made. However, the configuration name scoring issue was not identified until the study was in full progress. There was no way to determine which subjects in Sessions 3-8 changed the name of either the PC or the Server; thus, data were lost.

A total of 65 subjects completed the study during Sessions 3-8. These first 65 subjects were considered the first half of the sample in this study because starting with Session 9 and throughout the rest of the study subjects were instructed not to change the configuration name from "Client" or "Web Server." The media console in the lab was used to show subjects where on PT the configuration name was located on both the PC and the server. They were then told that if they changed either of these names, they would lose all points for that part of the activity. There were 81 subjects in Sessions 9-16 who received this additional instruction. Yet, for this second half of the sample there were other PT-related issues that resulted in lost data.

A second major issue with the PT activity scores were crashes, logouts, and computer freezes. For the first half of the sample, two PT activity scores were lost due to crashes. Thus, there were 63 PT activity scores for the first half. For the second half of the sample, 11 PT activity scores were lost. Six were lost due to PT crashes, three due to student logouts, and two due to computer freezes. Thus, there was a total of 70 PT scores for the second half.

A test for differences in scores on the PT activity was conducted between the two sample halves (63 vs. 70), and a significant difference was found, $F(1, 131) = 10.701$, $p < .001$, Cohen's f = .40 (large). The first half's PT activity scores ($M = .61$, $SD = .22$) were significantly lower than the second half's scores (M = .72, SD = .16). The scores for the first half of the sample are unreliable and cannot be used. Thus, only the 70 PT activity scores from the second half of the sample were included in the data anlaysis.

For the retention test, no scores were lost due to crashes or scoring issues. Thus, there were 65 subjects in the first half of the sample and 81 subjects in the second half. Univariate analyses were conducted to test for significant differences between the sample halves (65 vs. 81). No significant difference was found between the first and second halves of the sample on the retention test, $F(1, 134) = .013$, $p > .05$. However, we threw out the first half of the sample's retention scores due to the aformentioned scoring issue. Results from statisitcal tests on retention using full sample (N = 146) are presented in Appendix K. However, the results discussed in this chapter include 81 retention scores from the second half of the sample. The 11 PT crashes did not affect the pretest or retention scores.

Ultimately, the PT activity scoring issue resulted in a large portion of the sample being lost. The reduced sample size thus resulted in only 10-15 subjects per cell and therefore issues of power, which is discussed later in this chapter.

Scoring

The dependent variables under investigation were the retention post-test (scores ranging from 0-14) and the PT activity (scores ranging from 0-9). A pretest measured prior computer networking knowledge (scores ranging from 0-10). Percent correct scores were calculated for all three measures. PT's assessment tree system automatically scored the subjects' fixed networks. The assessment tree compared the network the subject tried to fix with an Answer Model, which was a properly configured network. PT then computed scores for the activity. Additionally, there were 17 five-point Likert scale self-report items used to assess affective experiences and as treatment fidelity checks. A significance level of .05 was applied for all statistical tests.

Overview

Because there was a significant correlation between the pretest and the retention test, an analysis of covariance (ANCOVA) was carried out for the retention test using the pretest as a covariate. However, there was no correlation between the pretest and the PT activity transfer test. Thus, an analysis of variance (ANOVA) was carried out for the PT activity transfer test. Both the analyses tested differences between the two factors: modality (text-only, voice-only, and voice+text) and restriction (restricted PT interface vs. non-restricted PT interface). Table 4.1 summarizes the mean percent correct scores (and standard deviations) for the six groups on the pretest, retention test, and the PT Activity transfer test.

Table 4.1 – Mean and Adjusted Mean percent correct scores and standard deviations for dependent measures.

	Text-Only						**Voice-Only**						**Voice+Text**					
	Restricted			**Unrestricted**			**Restricted**			**Unrestricted**			**Restricted**			**Unrestricted**		
	M	*SD*	*N*	*M*	*SD*	*N*	*M*	*SD*	*N*	*M*	*SD*	*N*	*M*	*SD*	*N*	*M*	*SD*	*N*
Prior Knowledge	.53	.21	12	.49	.19	15	.59	.16	14	.56	.14	14	.46	.16	14	.59	.21	12
Retention	.61a	.22	12	.67a	.17	15	.58a	.11	14	.60a	.14	14	.64a	.17	14	.57a	.17	12
PT Activity Transfer	.72	.16	10	.66	.18	11	.64	.15	13	.84	.08	11	.78	.14	13	.70	.19	12

Note: "a" means adjusted means using the pretest covariate. The potential range of scores was 0-10 for the pretest, 0-14 for the retention test, and 0-9 for the PT Activity.

Pretest

Assumptions

Independence of Observation

Subjects were randomly assigned to conditions in order of appearance to each session. Each computer station was prepared exactly the same for each student other than voice conditions having a headset. Students did not interact with each other at any point during the data collection sessions. Thus, observations were independent.

Normality

Figure 4.1 below shows the histogram for the pretest and a relatively normal distribution. Kolmogorov-Smirnov tests were conducted for the pretest (Kolmogorov-Smirnov Z = .175, $p < .05$). Table 4.2 below shows the results for each group. The tests results show a significant non-normal distribution for the pretest, and the group tests showed that Group 3 and Group 5 were significantly non-normal. However, Stevens (2002) suggested that ANOVA is relatively robust to normality violations.

Figure 4.1 – Histogram for Pretest.

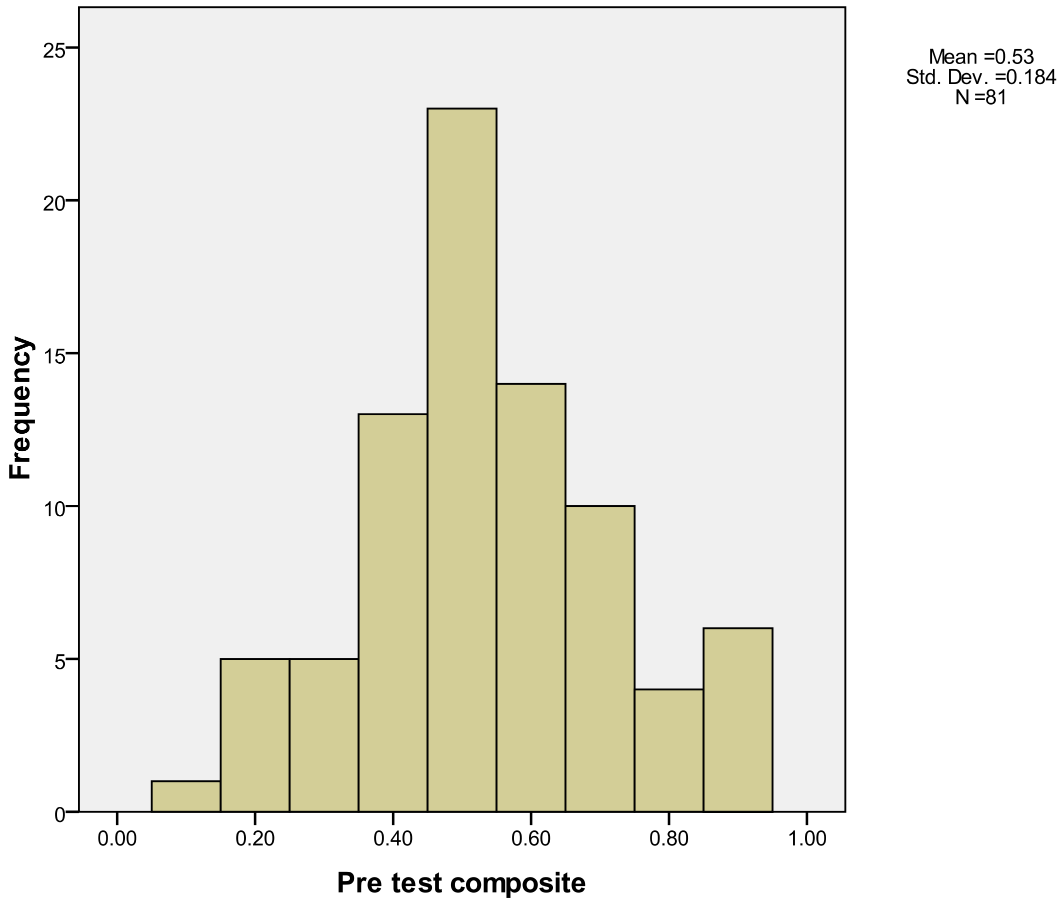


Table 4.2 – Kolmogorov-Smirnov Z for the Pretest by Group

Groups	Pretest
Group 1 (TR)	Z = .186, df = 10, $p > .05$
Group 2 (VR)	Z = .138, df = 13, $p > .05$
Group 3 (VTR)	Z = .323, df = 13, $p < .01$**
Group 4 (TU)	Z = .211, df = 11, $p > .05$
Group 5 (VU)	Z = .279, df = 11, $p < .05$*
Group 6 (VTU)	Z = .171, df = 12, $p > .05$

Equality of Variance

To test for equality of error variance for the pretest, Levene's test was carried out for the pretest. No significant difference was found $F(5, 75) = 1.157$, $p > .05$. Thus, the null hypothesis that the error variance of the dependent variable (retention test) is equal across groups is not rejected.

Pretest reliability

The internal consistency of the measures was low given the small number of items. The pretest had 10 items with a Cronbach's alpha of .37. Since the reliability of the measure is below .7, the results from analyses using this measure must be interpretted with caution. This is discussed more in the next chapter.

ANOVA Results for Pretest

The mean score on the prior knowledge test was .53 ($SD = .18$), indicating that subjects had some prior knowledge of computer networking. A 2 (restricted vs. unrestricted) x 3 (text, voice, voice+text) ANOVA was carried out to test for an interaction between the factors on the pretest scores; however, no significant interaction was found, $F(2, 75) = 1.707$, $p > .05$, Cohen's f = .21 (small). No significant main effects were found for either modality, $F(2, 75) = .976$, $p > .05$, Cohen's f = .16 (small) or restriction, $F(1, 75) = .444$, $p > .05$, Cohen's f = .07 (small).

Retention Post-test

Prior knowledge has been found to correlate with learning in computer-based environments (Kalyuga et al., 2003). Thus, the pretest of computer networking prior knowledge was used as a covariate in an ANCOVA that tested for differences between modality (text-only, voice-only, and voice+text) and restriction (restricted PT interface vs. non-restricted PT interface) on the retention test scores.

Assumptions

Independence of Observation

Subjects were randomly assigned to conditions; the computer stations were prepared exactly the same; and students did not interact with each other. Thus, observations were independent.

Normality

Figure 4.2 shows a histogram for the retention test, which is slightly negatively skewed; however, Stevens (2002) has suggested that ANOVA is relatively robust to normality violations. To test for the Normality assumption, the Kolmogorov-Smirnov test was conducted for the retention test (Kolmogorov-Smirnov Z = .136, $p < .05$). Table 4.3 lists the Kolmogorov-Smirnov tests for each group. Group 2 was significantly non-normal.

Figure 4.2 – Histogram for Retention post-test

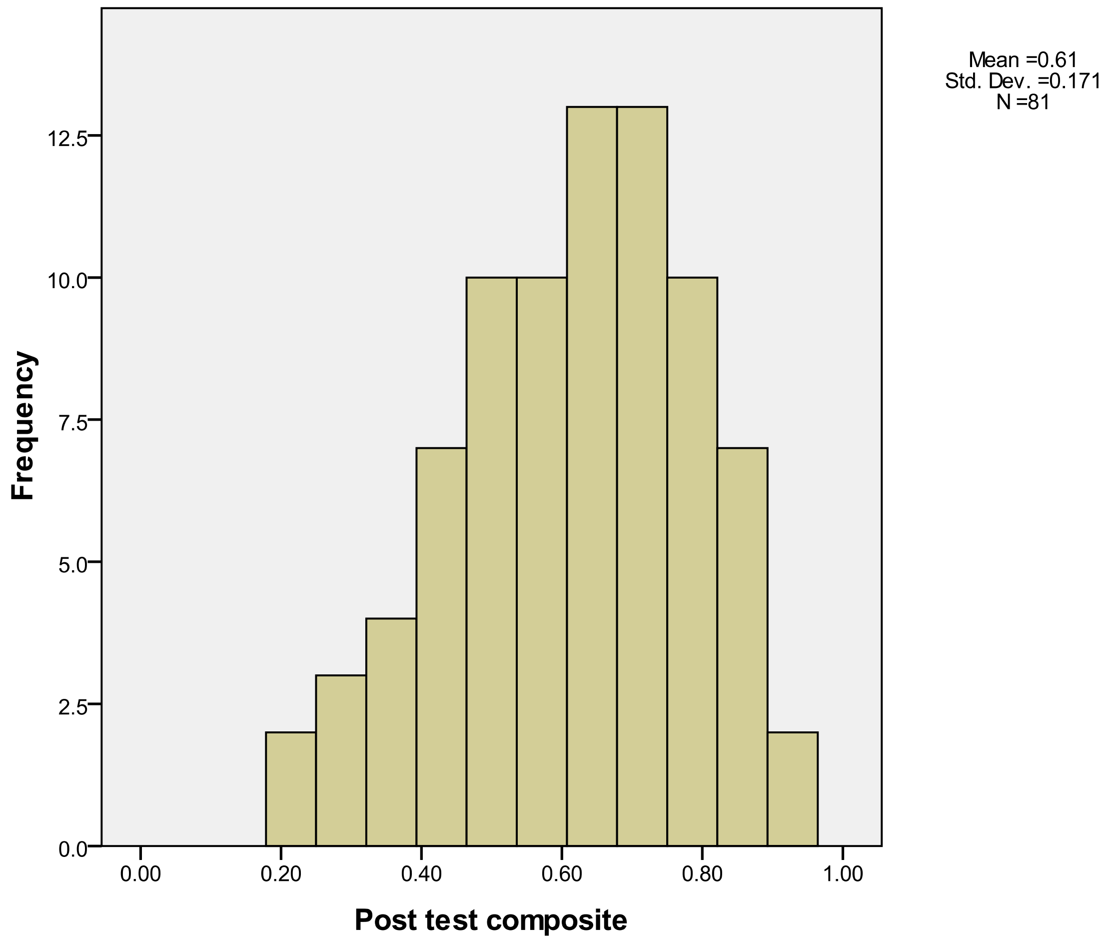


Table 4.3 – Kolmogorov-Smirnov Z for the Retention test by Group

Groups	Retention test
Group 1 (TR)	Z = .205, df = 10, $p > .05$
Group 2 (VR)	Z = .246, df = 13, $p < .05$*
Group 3 (VTR)	Z = .182, df = 13, $p > .05$
Group 4 (TU)	Z = .125, df = 11, $p > .05$
Group 5 (VU)	Z = .185, df = 11, $p > .05$
Group 6 (VTU)	Z = .154, df = 12, $p > .05$

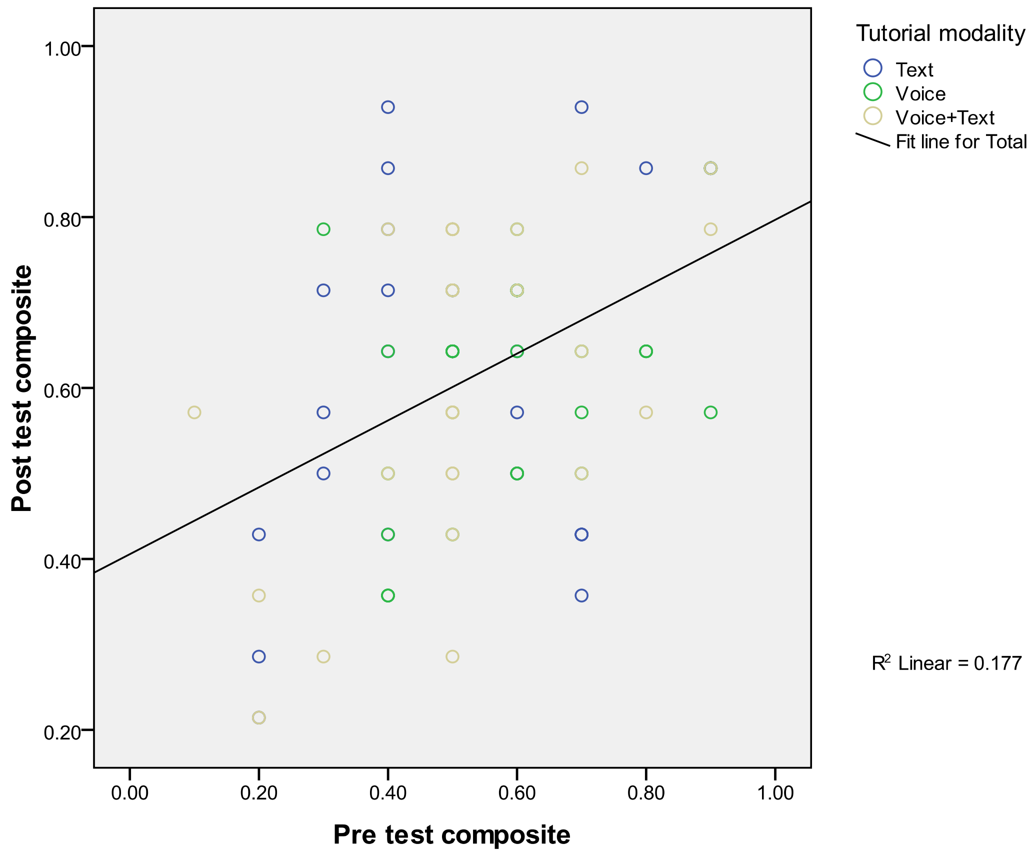
Equality of Variance

To test for equality of error variance, Levene's test was carried out for the 2 (restricted vs. unrestricted) x 3 (text, voice, voice+text) ANCOVA on the retention test. No significant difference was found for the retention test, $F(5, 75) = 1.043$, $p > .05$. Thus, the null hypothesis that the error variance of the dependent variable (retention test) is equal across groups is not rejected.

Linearity

A linear relationship was found between the covariate, prior knowledge of computer networking, and the retention dependent variable. A Pearson Product Moment correlation was calculated for the pretest of computer networking knowledge and the tutorial retention post-test. A significant correlation was found between the retention test and pretest, $r = .42$, $p < .000$. Figure 4.3 below illustrates the scatter plot for the pretest and retention post-test. Appendix L shows the scatter plots for pretest by post-test by each Group.

Figure 4.3 – Scatterplot for the pretest and retention post-test.



Equality of Regression Slopes

To test for equality of regression slopes, tests for an interaction of the joint effects between each factor and the covariate were conducted. Using SPSS, an ANOVA was run with a model that included all main effects of the factors, the covariate, and the interaction of the covariate with the factors. No significant interaction was found between the pretest composite score and modality, $F(2, 71) = .755$, $p > .05$. And, no significant interaction was found between between the pretest and restriction, $F(1, 71) = .112$, $p > .05$. In addition, no significant three way interaction was found between the covariate and the two factors, $F(2, 71) = 1.034$, $p > .05$. The interaction effect was non-significant; thus, the regression slopes and regression plane were homogeneous. The assumption was not violated.

Retention reliability

The retention test had 14 items with a Cronbach's alpha of.49. Once again, the low reliability is due to a small number of items, and since the alpha value is under .70 results most be interpretted with caution.

ANCOVA Results for Retention Test

An ANCOVA was carried out using only the second half of the sample (N = 81). No significant main effect was found for modality on retention, $F(2, 74) = .679$, $p > .05$, Cohen's f = .13 (small). There was no significant main effect of restriction, $F(2, 74) = .15$, $p > .05$, Cohen's f = .04 (small). No significant interaction was found for modality and restriction, $F(1, 74) = 1.141$, $p > .05$, Cohen's f = .17 (small).

ANOVA Results for Retention Test

Due to the low reliability of the pretest (Cronbach's alpha = .37), caution was used during statisical analyses. Additional analyses were conducted without using the pretest as a covariate to control for prior knowledge. A 2 (restricted vs. unrestricted) x 3 (text, voice, voice+text) ANOVA was conducted on the retention test. No significant differences were found.

Transfer Test: Packet Tracer Activity

Assumptions

Independence of Observation

Subjects were randomly assigned to conditions; the computer stations were prepared exactly the same; and students did not interact with each other. Thus, observations were independent.

Normality

The PT activity was designed to be a completion test in which subjects get a point for each correctly completed task while fixing PT. Figure 4.4 shows the histogram for the PT activity transfer scores. To test for the Normality assumption, the Kolmogorov-Smirnov test was conducted for the PT activity score transfer test, Kolmogorov-Smirnov $Z = .182$, $p < .05$. Table 4.4 shows Kolmogorov-Smirnov test results for each group. Groups 2,5, and 6 were significantly non-normal. Additionally, the skewness was -.804 and the kurtosis was .031. While these results indicate that the PT activity scores are non-normal, Stevens (2002) suggested that ANOVA is relatively robust to normality violations.

Figure 4.4 – Histogram for PT Activity

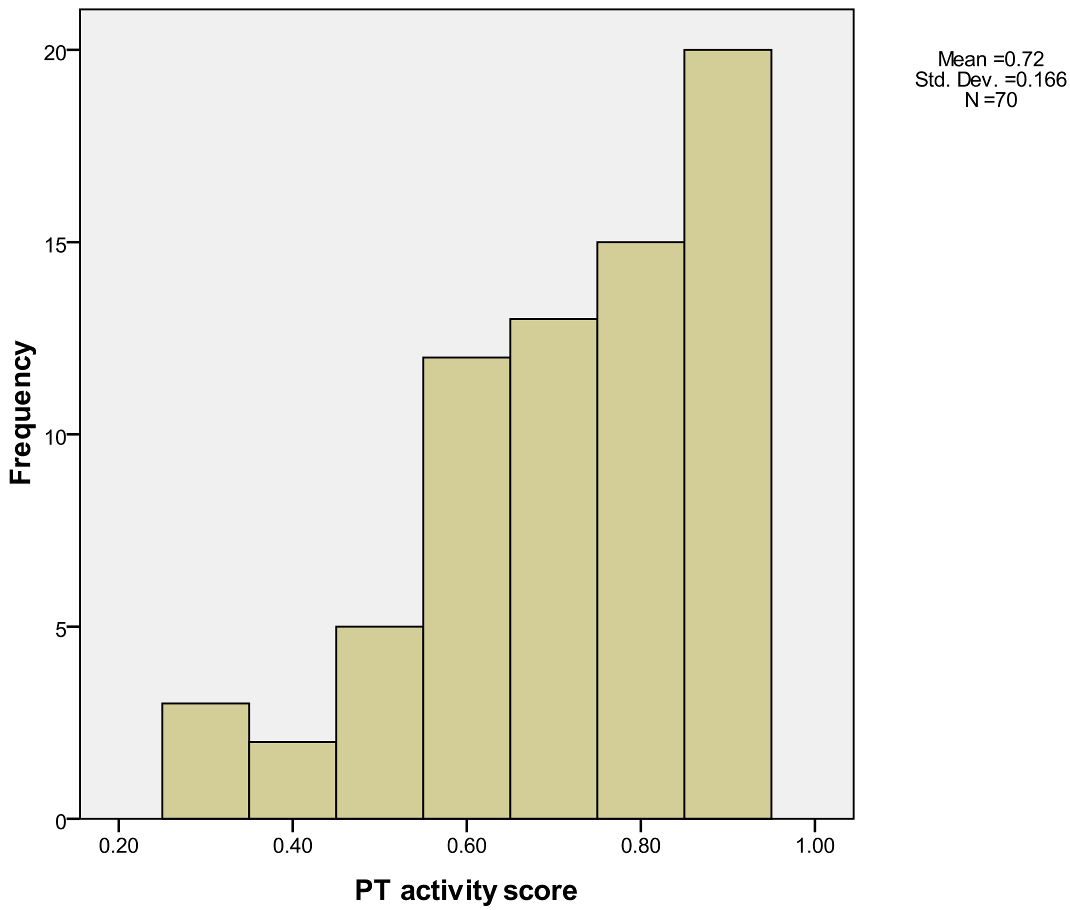


Table 4.4 – Kolmogorov-Smirnov Z for the PT Activity by Group

Groups	PT Activity Transfer Test
Group 1 (TR)	Z = .257, df = 10, *p* > .05
Group 2 (VR)	Z = .245, df = 13, *p* < .05*
Group 3 (VTR)	Z = .218, df = 13, *p* > .05
Group 4 (TU)	Z = .132, df = 11, *p* > .05
Group 5 (VU)	Z = .330, df = 11, *p* < .01*
Group 6 (VTU)	Z = .283, df = 12, *p* < .05*

Equality of Variance

To test for equality of error variance, Levene's test was carried out for the 2 x 3 ANOVA on PT activity transfer test. No significant difference was found for the PT activity, $F(5, 64) = 1.465$, $p > .05$. The null hypothesis that the error variance of the dependent variable (transfer test) is equal across groups is not rejected.

PT Activity Transfer Test Reliablity

The PT activity score was a composite score determined by the number of tasks correctly completed; thus, a reliablity score could not be calculated.

ANOVA Results for PT Activity Transfer Test

For the PT activity transfer test (N = 70), no significant main effect was found for modality, $F(1, 64) = .632$, $p > .05$, Cohen's f = .13 (small). There was no significant main effect of restriction, $F(1, 64) = .326$, $p > .05$, Cohen's f = 0.10 (small). A significant interaction was found between modality and restriction on the PT Activity, $F(2, 64) = 5.619$, $p < .01$, Cohen's f = .40 (large). Table 4.5 lists the results of the 2 x 3 ANOVA on the PT Activity measure.

Table 4.5 – ANOVA Source Table: PT activty measure

Source	**Sum of Squares**	**df**	**Mean Square**	**F**	**Sig.**	**Cohen's f^2**
Restrict	.008	1	.008	.326	.570	.10
Modality	.031	2	.016	.632	.535	.13
Interaction	.278	2	.139	5.619	.006	.40
Error	1.581	64	.025	-----		

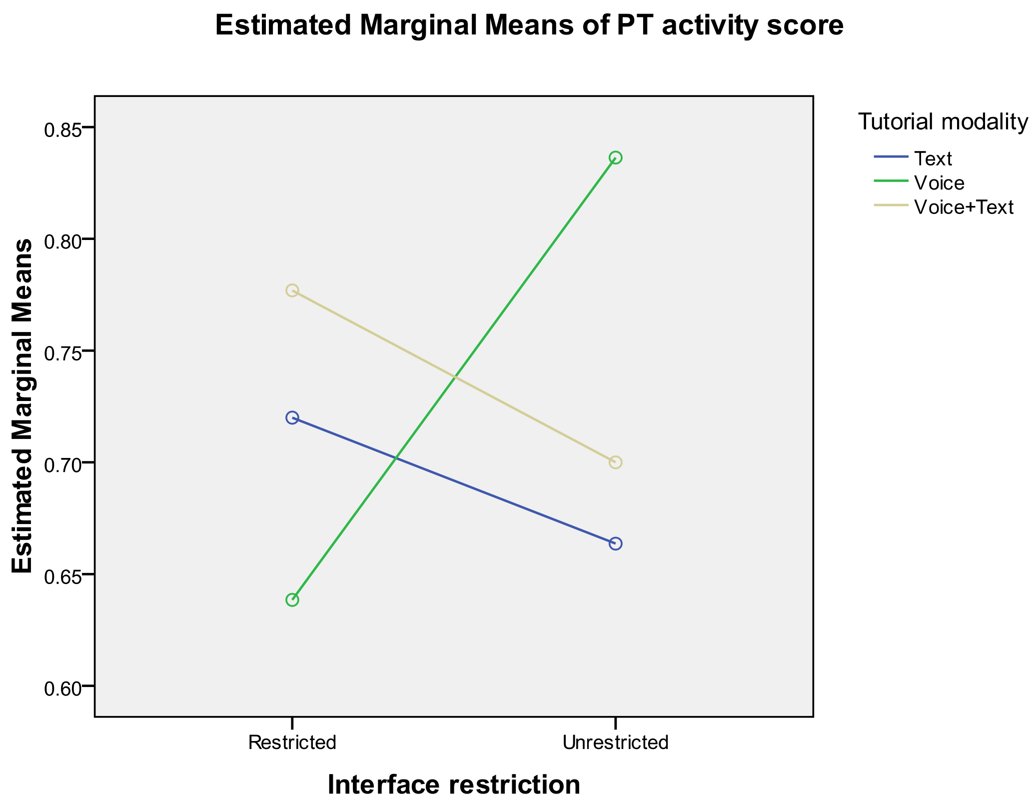
Tests of the simple effect of modality within each of the two restriction conditions were conducted to follow up the interaction effect. There was a significant simple effect within the unrestricted condition, $F(2, 31) = 3.502$, $p < .05$. A Tukey's HSD post hoc test showed that when PT was unrestricted those who received the voice condition scored significantly higher ($M = .836$, $SD = .08$) than those under the text-only condition ($M = .663$, $SD = .12$), $p < .05$.

The simple effect of modality within the restricted condition was not significant, $F(1, 64) = .326$, $p > .05$. Table 4.5 lists the means and standard deviations for the interaction; and Figure 4.5 illustrates the two-way interaction between modality and restriction on the PT activity measure.

Table 4.6 – Means for Two-Way Interaction between Modality and Restriction on PT Activity

Means, Standard Deviations, and Cell sizes				
Tutorial modality	**Interface restriction**	**M**	**SD**	**N**
Restricted	Text	.720	.16	10
	Voice	.638	.15	13
	Voice+Text	.776	.14	13
	Total	.711	.15	36
Unrestricted	Text	.663	.12	11
	Voice	.836	.08	11
	Voice+Text	.700	.19	12
	Total	.732	.17	34
Total	Text	.690	.16	21
	Voice	.729	.15	24
	Voice+Text	.740	.16	25
	Total	.721	.16	70

Figure 4.5 – Two-Way Interaction between Modality and Restriction on PT Activity



Lack of Power

Due to the PT activity scoring issue the sample was greatly reduced in size thereby creating a power issue. Observed power for the the PT activity interaction was .843; however, this was the only significant finding. All other power estimates were below .35. Power for the PT activity main effect of modality was .151 and restriction was .087. For the retention test, observed power was .078 for the main effect of modality, .069 for the main effect of restriction, and .076 for the interaction. For the pretest, observered power was .214 for the main effect of modality, .101 for restriction, and .348 for the interaction. Examining the observed power for the statistical tests provides evidence that the PT activity scoring issue had a large impact on the study in terms of reducing the sample and thereby lowering power to detect significant differences caused by the factors.

Results & Research Questions

The following section summarizes the results in the context of the research questions:

RQ1. Given research on the modality effect, how does the "My 1st PT Lab" tutorial instructional delivery modality (text-only, voice-only, and voice+text) affect performance on dependent measures?

No main effect of modality was found on the retention test of tutorial knowledge. This null finding was due to a small effect size (Cohen's f = .13) rather than insufficient sample size (N = 26-28). Similarly, no main effect for modality was found on the PT activity transfer test due to the small effect size (Cohen's f = .13) rather than sample size (N = 21-25).

What effect does PT's interface being restricted have on dependent measures?

No main effect of restriction was found on the retention test. This null finding was due to a small effect size (Cohen's f = .04) rather than insufficient sample size (N = 40-41). No main effect of restriction was found on the PT activity transfer test due to a small effect size (Cohen's f = .10) rather than insufficient sample size (N = 34-36).

Is there an interaction between modality and interface restriction?

No interaction between modality and restriction was found on the retention test once again due to the relatively small effect size (Cohen's f = .17) and a small sample size (12-15). However, a significant two-way interaction was found between modality and restriction on the PT activity transfer test due to a large effect size (Cohen's f = .40), despite a small sample size (10-13). Follow up tests showed that voice-unrestricted group? scored significantly higher than text-unrestricted on the PT activity. In other words, adding voice to the tutorial and taking away extraneous text led to increases in transfer from the tutorial to the PT activity. It is critical to note that this significant effect occurred only when PT's interface was unrestricted. This result can be explained by Mayer's cognitive theory of multimedia learning (2001, 2005) and will be discussed further in the next chapter.

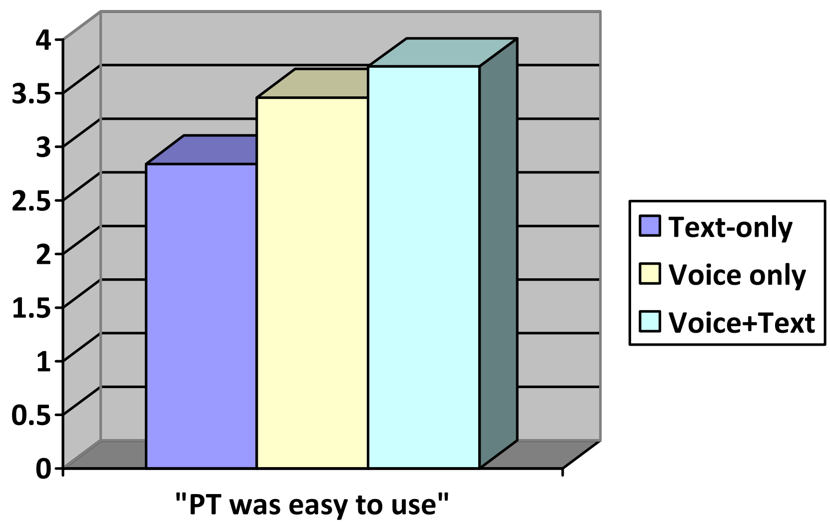
Self-report data

Self-report items were intended to measure four aspects of the subject's experience: cognitive load (extraneous, intrinsic, and germane) and affective response to the study. The following section discusses results for these four areas. ANOVAs were carried out to test for differences of modality and restriction as well as an interaction. However, because the self-report items may be viewed as ordinal measures, non-parametric Kruskal-Wallis H tests were conducted as follow ups for any items found to have significant differences among conditions to control for Type I error inflation.

Extraneous cognitive load items

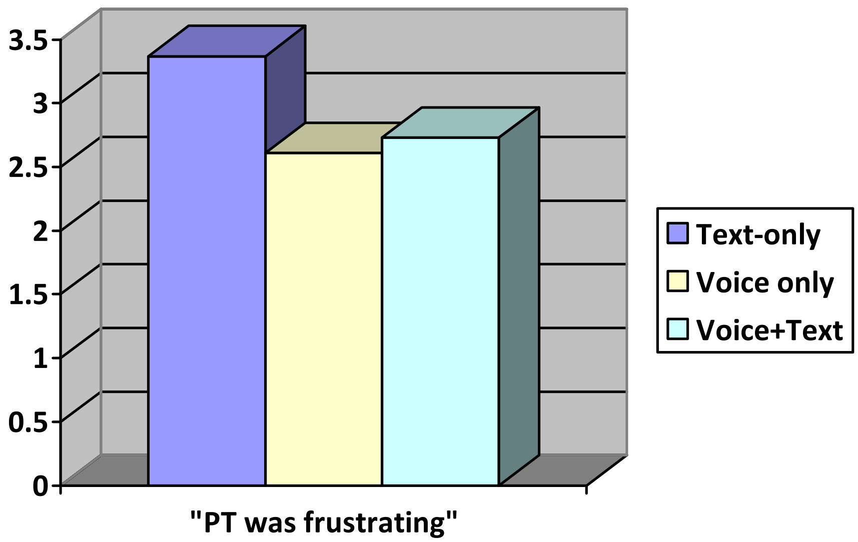
Two extraneous cognitive load items were found to have significant differences. A significant main effect of modality was found for the self-report item "Packet Tracer's interface was easy," $F(2, 75) = 5.080$, $p < .01$, Cohen's F = .35. Tukey's HSD post hoc tests revealed that both the voice (M = 3.46, SD = .83) and voice+text ($M = 3.75$, $SD = 1.07$) groups perceived PT to be easier to use than did the text-only group ($M = 2.84$, $SD = 1.27$), $p < .05$.However, there was only a marginally significantly difference, $p = .066$, between voice and text. Thus, having voice accompany the text in the tutorial positively affected how subjects perceived PT in terms of ease of use (see Figure 4.6 below). The Kruskal-Wallis H test confirmed the parametric results, X^2 (2, n = 81) = 8.699. $p < .05$.

Figure 4.6 – Results for Self-Report item “PT’s interface (visual display) was easy to use.”



A significant main effect of modality was also found for “Packet Tracer was frustrating,” $F(2, 75) = 4.060$, $p = .021$, Cohen’s F = .32. Tukey’s HSD post hoc tests revealed that the text group ($M = 3.37$, $SD = 1.18$) found PT significantly more frustrating than the voice group ($M = 2.61$, $SD = 1.03$) and the voice+text group ($M = 2.73$, $SD = .96$), $p < .05$. Therefore, the text-based tutorial negatively affected how students perceived PT in terms of frustration compared to the voice groups. The Kruskal-Wallis H test was significant, X^2 (2, n = 81) = 6.892. $p < .05$, confirming the parametric findings. Figure 4.7 illustrates the differences in the means for the item.

Figure 4.7 – Results for Self-Report item “PT was frustrating”



Non-significant extraneous cognitive load items

Descriptive statistics from the non-significant items shed light on subjects' experiences. Two items that all subjects completed were “Packet Tracer was fun” and “Packet Tracer's interface was overwhelming.” Figures 4.8 and 4.9 below chart the means for the six groups on the two items. While the responses on the items are not significantly different, the results are consistent with the two aforementioned extraneous cognitive load items. According to these two items, subjects in the text-only tutorial had less fun using PT compared to the voice groups, and subjects in the text-only tutorial perceived PT's interface to be overwhelming compared to voice groups. However, this was only notable when PT's interface was unrestricted.

Figure 4.8 – Results for Self-Report item “Packet Tracer was fun.”

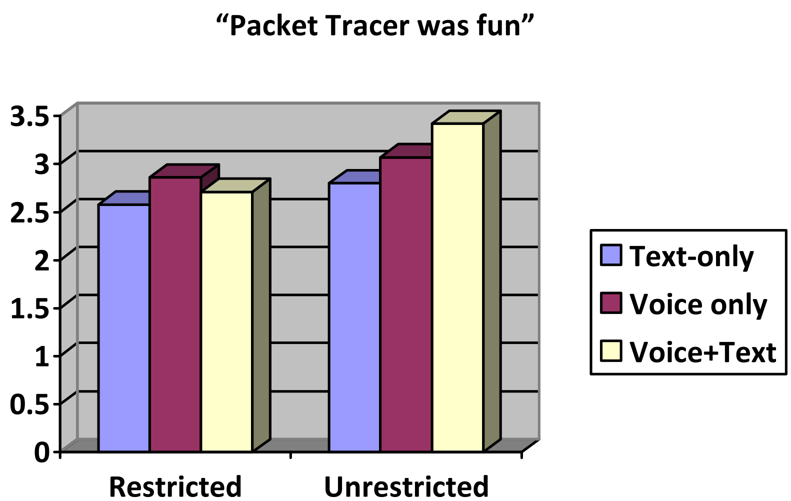
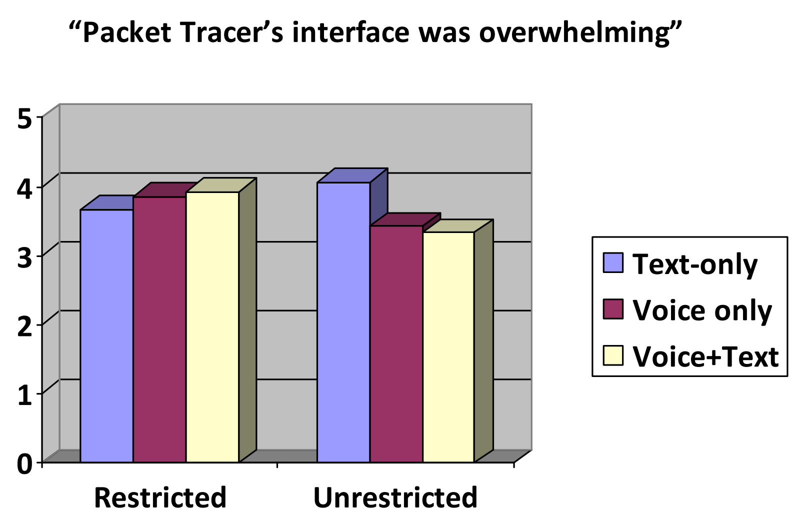


Figure 4.9 – Results for Self-Report item “Packet Tracer’s interface was overwhelming.”



Restricted-only items

Subjects in the restricted conditions received two additional extraneous cognitive load items. These included: “The ‘Locked’ messages distracted me” and “Not being able to access a restricted part of Packet Tracer made me curious about why I was being restricted." The unrestricted groups did not answer these items because they only received a "Locked" message if they clicked a check results button. Thus, not all of the subjects experienced a "Locked" message. Figures 4.10 and 4.11 below chart the means for the six groups on the two items. The results show that for subjects in the restricted groups, those who were in the text group considered the "Locked" messages less distracting compared to subjects in the voice groups, perhaps due to the text group having to read text popup windows throughout the tutorial. Second, subjects in the text group reported that the restricted parts of the interface made them more curious compared to the voice groups.

Figure 4.10 – Results for Self-Report item "Locked messages distracted me."

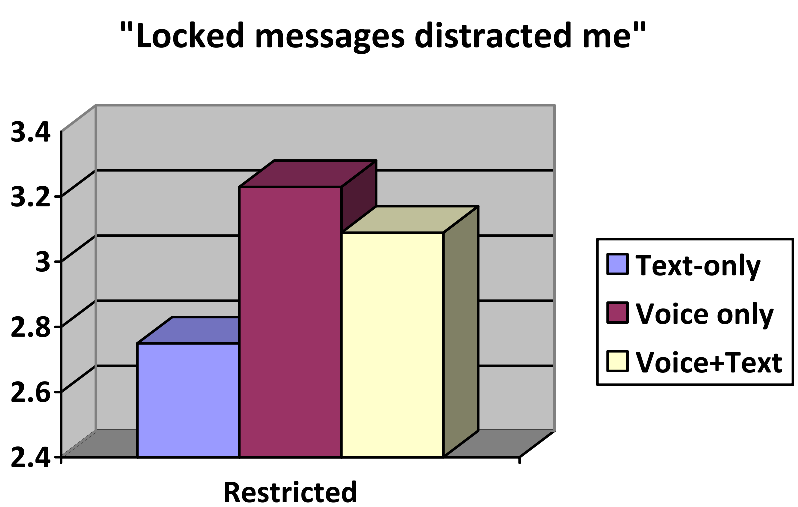
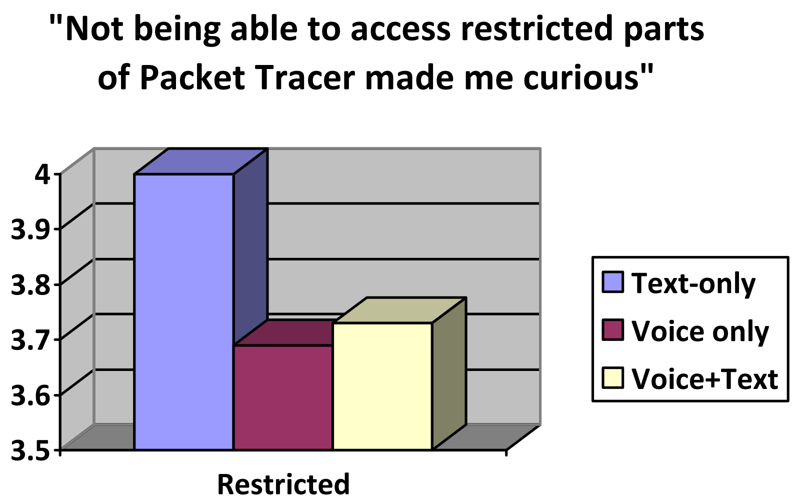


Figure 4.11 – Results for Self-Report item "Not being able to access restricted parts of Packet Tracer made me curious."



Intrinsic cognitive load items

ANOVAs on these items revealed no significant effects of modality or restriction and no interaction. However, charts for the four items are shown below (see Figures 4.12 – 4.16). An interesting observation is that the voice-unrestricted group reported that PT required less mental effort and was less complex compared to ratings for all five other groups (see Figures 4.13 and 4.15). A second observation is that the text group, regardless of restricted or unrestricted, rated the tutorial as requiring more mental effort and being more complex compared to the voice groups, which is logical since the text groups had to read the narration while the voice groups (voice and voice+text) could listen to the narration.

Figure 4.12 – Results for Self-Report item “Mental Effort during tutorial.”

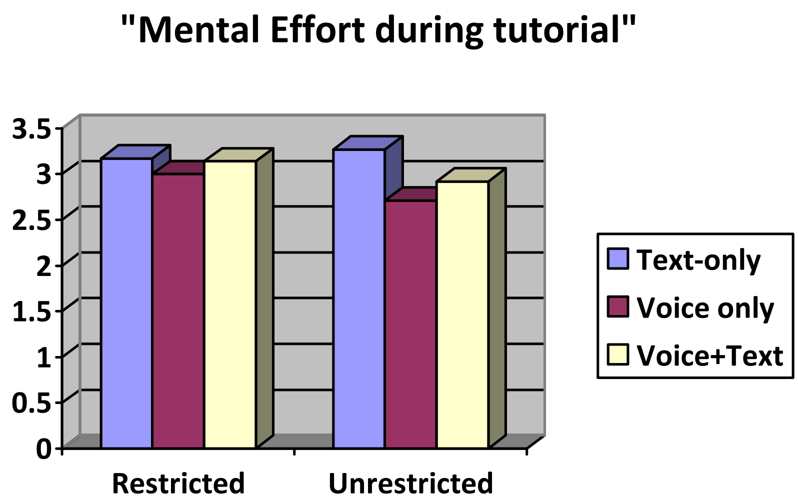


Figure 4.13 – Results for Self-Report item “Mental Effort using PT.”

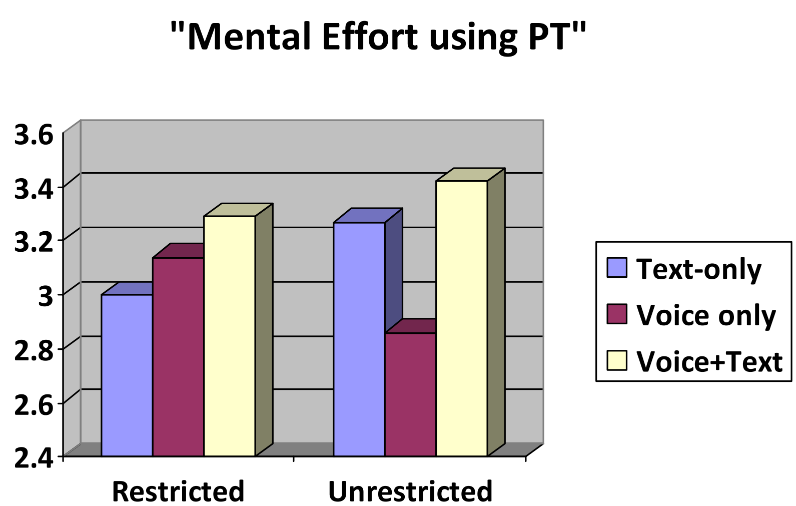


Figure 4.14 – Results for Self-Report item “Tutorial’s level of complexity.”

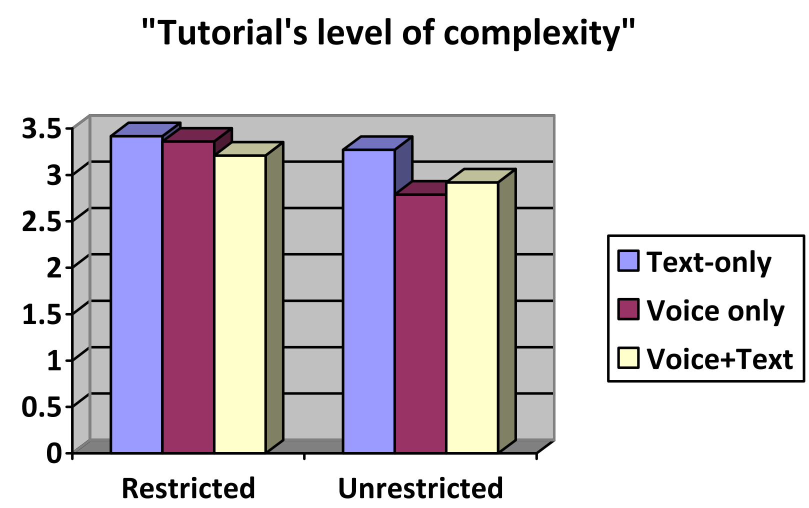
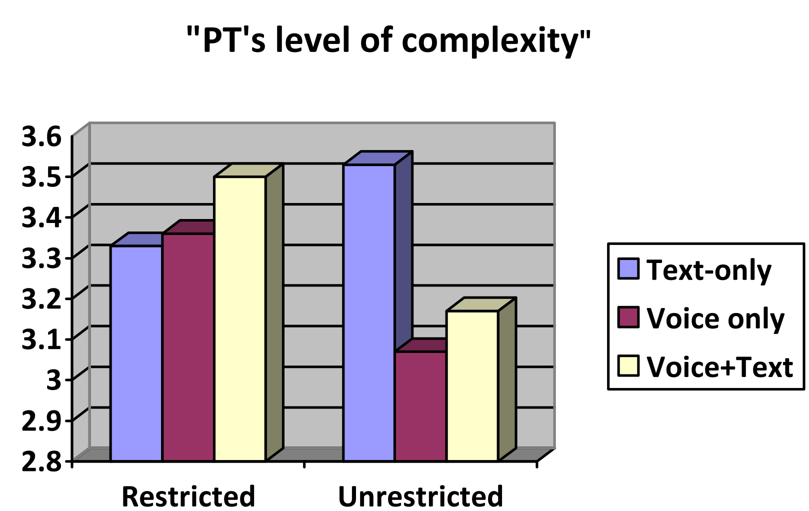


Figure 4.15 – Results for Self-Report item “Mental Effort using PT.”



Germane cognitive load items

ANOVAs on germane cognitive load items revealed no significant effects of modality or restriction and no interaction. However, charts for the four items are shown below (see Figures 4.16 – 4.18). The difficulty ratings did not show much in terms of differences between groups' ratings understanding the tutorial, remembering the tutorial, and using PT. However, text groups reported having slightly more difficulty remembering information from the tutorial.

Figure 4.16 – Results for Self-Report item "Difficulty understanding the tutorial."

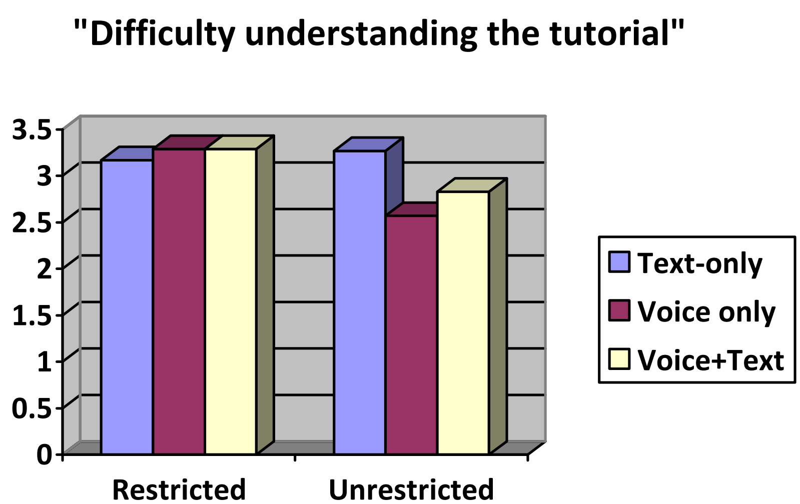


Figure 4.17 – Results for Self-Report item "Difficulty remembering information from the tutorial."

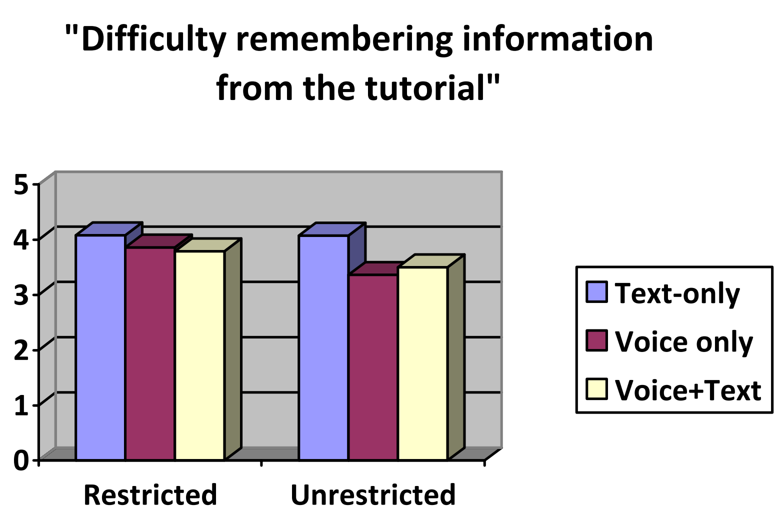
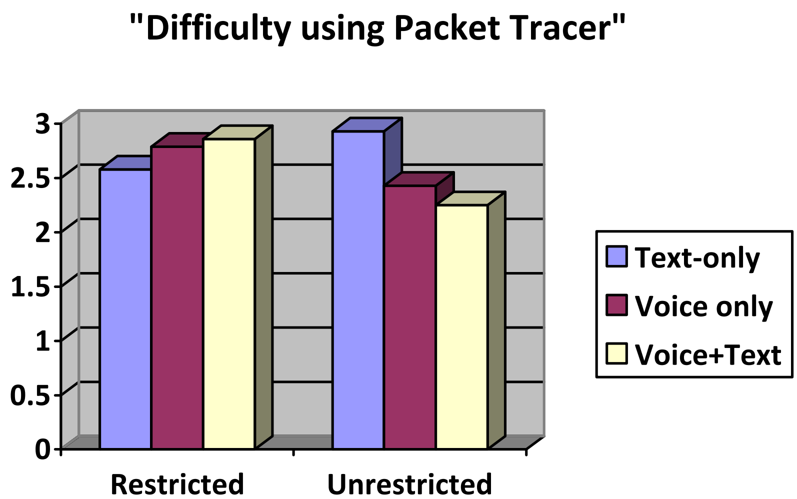


Figure 4.18 – Results for Self-Report item “Difficulty using Packet Tracer.”



Affective self-report items

ANOVAs on germane cognitive load items revealed no significant effects of modality or restriction and no interaction. Charts for the three items are shown below (see Figures 4.19 – 4.21). For the three affective items, compared to the five other groups, the voice+text-unrestricted group reported having higher eagerness to continue using PT, interest in the material, and amount of effort put into the study.

Figure 4.19 – Results for Self-Report item “Eagerness to continue using Packet Tracer.”

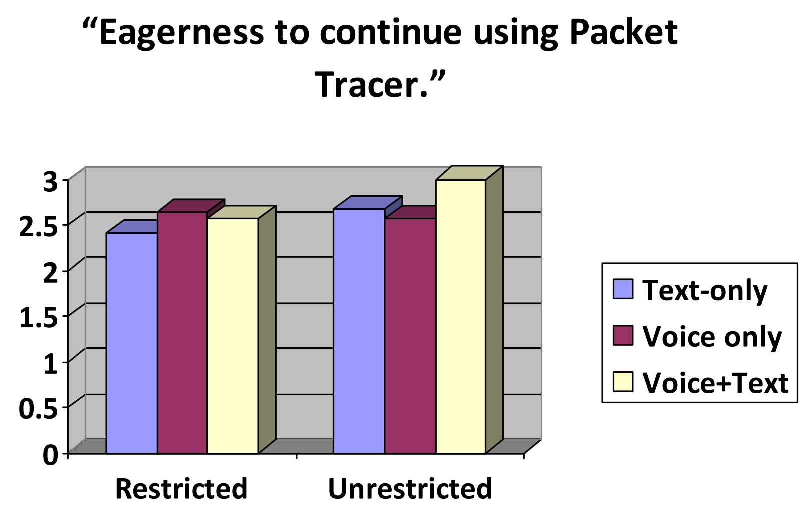


Figure 4.20 – Results for Self-Report item “How interesting is this material?”

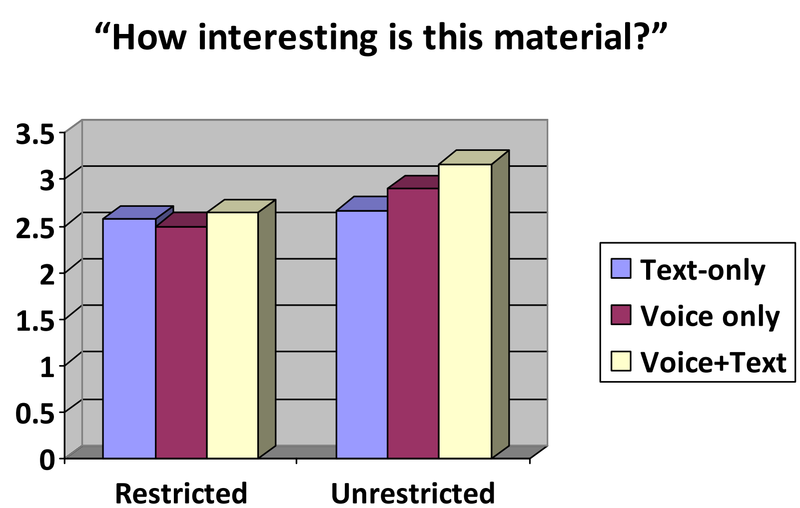
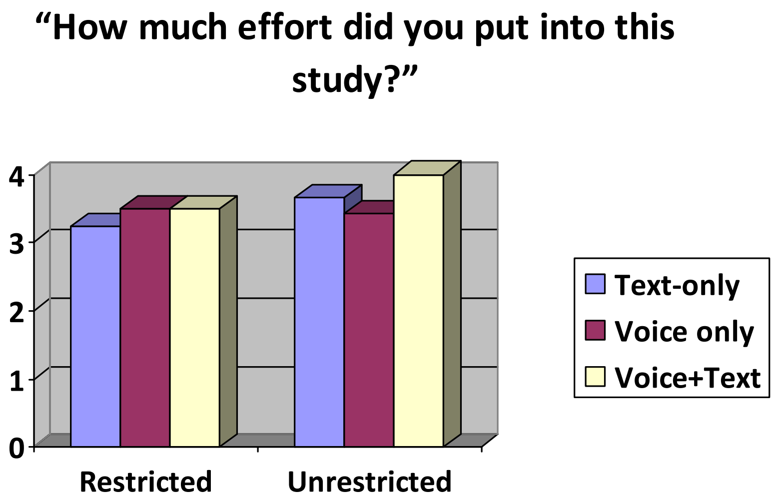


Figure 4.21 – Results for Self-Report item “How much effort did you put into this study?”



Self-report summary

An analysis of self-report data supports the significant interaction between modality and restriction for the transfer test. Two extraneous cognitive load self-report items were found to have statistical differences for the modality factor. The voice-only and voice+text groups perceived PT to be easier to use and less frustrating than the text-only group. This is consistent with the simple effect found that the voice-unrestricted group outperformed the text-unrestricted group on the PT activity. In terms of cognitive load, the text group reported having significantly higher extraneous load (frustration) during the PT activity compared to the voice groups.

The majority of self-report items did not reveal significant differences; however, examining their descriptive statistics showed that the voice-unrestricted group reported using less mental effort or intrinsic cognitive load during the PT activity. The text groups reported experiencing more intrinsic cognitive load during the tutorial compared to the voice groups. Results from the germane cognitive load measures of difficulty did not reveal many observations other than both text-restricted and text-unrestricted reported having slightly more difficulty remembering information from the tutorial.

Results from the affective measures showed that the voice+text-unrestricted group was the most interested, put in the most effort, and wanted to continue using PT compared to the other five groups. However, these differences were not significant possibly due to a lack of power due to sample size. Practical and theoretical implications from these results are discussed in the next chapter. These results support the advantage for the voice-only unrestricted condition over the text-only unrestricted condition.

Chapter 5 Discussion

This chapter includes a summary of the study, findings, limitations, both practical and theoretical implications, and recommendations for future research. The purpose of this study was twofold. One purpose was to evaluate what combination of multimedia is optimal for retention of tutorial knowledge and transfer of computer networking skills. The second purpose was to empirically test theoretical models of multimedia instructional delivery. Specifically, I wanted to see whether tutorial modality and restriction of a sim's functions affected performance due to differences in the demands on cognitive resources.

There were numerous issues that affected the sample size, such as the sim's embedded assessment system and the reliability of the pretest and retention test. These issues are common to educational technology research (Ross & Morrison, 1996); and, as technologies become more complex, the challenges become more sophisticated as well. For example, embedded assessments are powerful tools for measuring a student's ability to perform a task, such as fixing a broken computer network. However, such assessments are extremely sensitive to student error and generally graded using a dichotomous system. In other words, PT does not give partial credit. This chapter discusses the results from this study and how they inform existing literature as well as the challenges experienced while running a study using a complex simulation, a multimedia tutorial, and researcher-developed measures.

Summary

This study examined how to best deliver instructional content (text, voice, voice+text) in a multimedia tutorial that teaches Cisco Networking Academy students how to use PT, a computer networking sim that allows for exploration, construction, and assessment. As for the sim itself, the study examined whether restricting non-task-relevant features a user could use within PT would affect performance. Two measures of learning were used: a 14-item multiple choice test assessed retention of information from the tutorial and a computer networking solutions score ranging from zero to nine assessed transfer of information from the tutorial to PT. Very simply, the study examined the effects of tutorial modality and sim interface restriction on tutorial retention and transfer.

Theoretically, this study explored how two constructs may explain how tutorial modality and sim interface restriction affect retention and transfer. First, the cognitive theory of multimedia learning states that balancing the visual and verbal sensory input channels can facilitate learning (Mayer, 2001, 2005). The theory includes several principles that are examined in this study. The redundancy principle states that people learn more deeply from graphics and narration than from graphics, narration, and on-screen text. However, Mayer and Johnson (2008) revised the redundancy principle by varying the amount of on-screen text. A full text of what is being verbally narrated was redundant and led to extraneous processing; however, if only a few key words were presented in text next to the relevant parts of an animation, then learning increased due to directing the subjects' attention without inducing extraneous cognitive load.

In addition, one of the major tenets of Mayer's theory is the modality effect, which states that effectively utilizing the visual and auditory channels can increase working memory capacity and reduce cognitive load. This study extends both the redundancy principle and the modality effect by testing the effect of on-screen text, voice narration, and voice narration combined with on-screen text. For the voice+text condition, the text narration used in this study was redundant in that the whole narration was presented versus only presenting key words in the pop-up text boxes. Future research could connect this study with Mayer and Johnson's (2008) recent work by only presenting non-redundant text in the voice+text condition. This would entail only presenting key words and phrases, such as computer networking terminology or critical information needed to fix the computer network.

The second theoretical framework used in this study is cognitive load (Sweller, 1988; Sweller & Chandler, 1994). Cognitive load assumes that humans have a limited capacity to process information. Cognitive load theory distinguishes between three types of load: intrinsic, germane, and extraneous. Intrinsic load is the amount of cognitive resources used by completing a task. Germane load is the amount of resources used to create schemas and meaningful learning. Extraneous load is the amount of resources taken away from the former two due to distracting stimuli in the environment. Cognitive load has direct bearing on the design on multimedia tutorials and sims. Presenting too much information or unnecessary graphics on the screen can induce extraneous cognitive load within the learner's visual channel and thereby take attention away from selecting, organizing, and integrating relevant information (Baddeley, 1986; Mayer, 2001; Paivio, 1986). In this study, cognitive load was manipulated by both tutorial modality and by restricting access to parts of PT's user-interface. These two instructional manipulations (tutorial modality and sim interface restriction) were chosen because both were expected to have an effect on the learner's ability to use the PT.

Unique Aspects of the Study

This study is unique in four ways. First, there is no literature investigating how the modality of a tutorial that teaches how to use a complex sim interacts with whether the sim's interface is restricted or not. This type of examination is timely given the increasing dependence on complex sims for practicing skills. Second, Mayer and colleagues' research on the modality effect and redundancy principle has used researcher-created animations (Ginns, 2005). In contrast, this study used existing instructional materials that are currently being used by hundreds of thousands of Cisco Networking Academy students around the world. Third, although Mayer and colleagues frequently measure learning using both retention and transfer tests, their transfer tests often require subjects to write short answers to questions that apply concepts to new situations. In contrast, this study measured transfer by actual performance of the subject's ability to troubleshoot the network. Thus, this study is unique because it introduces a new type of transfer measure to modality research, e.g., embedded assessment within a sim. Fourth, this study examined the effects of restricting parts of a sim's interface on learning. There is to date no studies that have examined this manipulation. These four unique aspects of this study tie together recent advances in research on the cognitive theory of multimedia learning, cognitive load theory, instructional technology design, and assessment. Thus, the findings and lessons learned from this study are relevant and timely.

Findings

The findings in this study were limited due to a scoring issue which reduced sample size and therefore reduced statistical power to find differences caused by the manipulations of modality and restriction. Additionally, the findings in this study were limited due to researcher-developed measures with low reliability estimates. However, while there were issues, a significant interaction between modality and restriction was found for the PT activity. For the unrestricted groups, subjects who received the voice-only tutorial showed better transfer of knowledge to PT compared to subjects who received the text-only tutorial. In terms of the self-report data, subjects in the voice and voice+text conditions, regardless of whether their interface was restricted or not, perceived PT to be easier to use and less frustrating than subjects who received the text-only tutorial. Likewise, subjects in the text conditions rated PT as less easy to use and more frustrating than the voice and voice+text conditions. Thus, adding voice to their tutorial affected subjects' attitudes towards PT.

Research Questions

RQ1. Given research on the modality effect, how does the "My 1st PT Lab" tutorial instructional delivery modality (text-only, voice-only, and voice+text) affect performance on dependent measures?

It was hypothesized that voice-only would outperform both text-only and voice-plus-text due to the modality principle. This was partly supported for the transfer test (within the unrestricted conditions) but not the retention test. A modality by restriction interaction was found on the PT activity. When PT was unrestricted, receiving the tutorial with voice-only was significantly better than with text-only. This is consistent with the modality effect and the redundancy principle (Mayer, 2001, 2005). In particular, Mayer and colleagues have found that when presenting instructional content using a multimedia animation, voice narration rather than a text-based narration frequently leads to increased scores on transfer tests but not on retention tests. The cognitive theory of multimedia learning explains this phenomenon stating that voice and animation better utilizes human's sensory input channels than text and animation. The result is that in the voice condition, learners can more effectively attend to, select, and integrate relevant information. Further, according to cognitive load theory, voice and animation allow for germane processing which leads to schema construction and thus higher transfer scores compared to text and animation which require high levels of intrinsic processing due to splitting attention between the two visual sources of information (Harskamp et al., 2007; Sweller, 1988). Thus, the finding in this study that voice performed better than text on transfer scores is consistent with previous research on the modality effect.

RQ2. What effect does PT's interface being restricted have on dependent measures?

It was hypothesized that restricting the user's access to extraneous tools and objects on the PT interface would reduce distractions and lower the chance of getting lost within the sim. Thus, the potential for extraneous cognitive overload would be reduced and learners would have additional cognitive resources available for germane processing. However, no differences were found between the unrestricted and restricted conditions on either retention or transfer measures. After observing the differences between the two conditions, this hypothesis may have been overly simplistic and perhaps other factors affected how users reacted to PT being restricted.

The PT interface restriction manipulation was not as clean as necessary. The restricted condition had a third of the interface restricted; however, there were still numerous opportunities to get lost within the countless windows and options available in PT. Thus, the restricted condition was not restricted to the degree to which all possibilities for extraneous behavior were removed. Additionally, subjects in the restricted group received numerous messages stating that what they clicked on is "Locked." This message was presented with a loud bell noise. Perhaps, the restricted messages discouraged learners, similar to how receiving constant error messages increases frustration. However, according to the self-report data, restricted subjects did not report being more frustrated than unrestricted subjects. If the restricted messages provided helpful content, such as specific instructions, instead of "Locked" messages then perhaps this condition would have been more helpful.

The unrestricted subjects also received a "Locked" message if they clicked on the "Reset Activity" or "Check Results" buttons on the PT instructions window. Thus, the unrestricted conditions were still slightly restricted because subjects received a message with a bell if they clicked either button. One self-report item asked subjects if they received "Locked" messages. All subjects in the restricted conditions reported "yes"; however, a third of subjects in the unrestricted conditions also reported "yes". Thus, the degree to which these two treatments actually differed is an issue. Future research should use a cleaner manipulation of interface restriction, such as restricting all extraneous parts of the interface and having the unrestricted conditions receive zero "Locked" messages.

RQ3. Is there an interaction between modality and interface restriction?

Based on existing literature, it was hypothesized that the high load conditions would overload the visual channel thus resulting in poorer performance compared to the control and low cognitive load conditions (Mayer, 2001; Sweller, 1988). The hypothesized highest cognitive load conditions were expected to be text-only unrestricted and voice+text unrestricted. The lowest load condition was expected to be voice-only restricted, which was expected to score higher on the retention and transfer tests compared to other conditions that presented text on the screen. In addition, it was predicted that there would not be a significant interaction between modality and interface restriction. Contrary to predictions, a significant interaction between tutorial modality and interface restriction was found on the transfer measure. Subjects in the unrestricted condition who received voice-only narration scored higher on the PT activity compared to subjects who received a text-based narration. The content was exactly the same. The only difference between the voice-only tutorial and the text-only tutorial was that no text was presented during the voice-only tutorial. Presenting text in the tutorial affected subjects performance using PT; yet, this was found in the unrestricted conditions but not in the restricted conditions.

Explaining the Interaction

Within the unrestricted condition, from a practical sense, this finding can be explained by the modality effect. Subjects had to watch an animation of how to use PT while listening to or reading the narration. In the voice condition they were able to listen to narration and watch the animation. In contrast, the text condition split subjects' attention in the visual channel between the text pop-up windows and the animation. Split attention effects have been documented and are part of the foundation of both the cognitive theory of multimedia learning and cognitive load (Mayer, 2005; Paivio, 1986; Sweller, 1988).

Why this modality effect emerged in the unrestricted condition but not in the restricted condition is less clear. Perhaps rather than reducing extraneous cognitive load, the restricted conditions may have actually increased extraneous load because subjects received a high number of "Locked" messages with loud bell noises in their headphones.

An “OK” button had to be clicked to close a “Locked” message. Thus, the way the “Locked” messages were delivered could have produced higher extraneous cognitive load in the restricted conditions compared to the unrestricted conditions. This higher extraneous cognitive load in the restricted conditions therefore undermined the scaffolding that was intended to help the subject by reducing the amount of unnecessary buttons and functions on the sim’s interface.

On the other hand, the unrestricted conditions allowed subjects more freedom to explore. These subjects could click on and open anything in PT. They therefore had more opportunities for getting lost. Given the evidence from this study, it is clear that the modality of the tutorial affected subjects’ transfer of knowledge from the tutorial to PT. Another tenable explanation for the interaction is that subjects who received optimal instruction (voice-only tutorial) were perhaps not led to click task-irrelevant buttons; whereas, the text-only group may have gotten off-task more often due to less-than-optimal instruction that led to poor transfer performance.

Self report data help to explain the advantage of voice over text within the unrestricted condition on the PT activity transfer test. The voice and voice+text conditions rated PT as easier to use and less frustrating compared to the text condition. Again, consistent with the modality effect, subjects in the voice-only conditions were able to watch the animation of how to use PT while listening to the audio narration. In contrast, subjects who received text-only had to read the narration and watch the animation at the same time. This decreased performance on the transfer test likely due to the split attention effect. Self report data showed that the modality and interface restriction factors affected subjects’ attitudes towards PT.

There is no existing research on how restricting parts of a complex sim's interface can scaffold novices; thus, more research is needed on what are the most effective methods for scaffolding through interface restriction. One potential method could be to simply *grey out* the buttons and functions that are restricted rather than popping up a "Locked" message with a bell. Numerous software companies already use this method of not providing access to certain parts of a software depending upon how the software is being used. More research is needed to explore this possibility.

Lack of Findings for Cognitive Load

This study sought to build on previous research (Mayer, 2005; Mayer & Johnson, 2008) by utilizing Likert type self-report items to quantify three types of cognitive load: extraneous, intrinsic, and germane. Only two extraneous load items showed significant differences; however, there were a host of factors that might explain why the other cognitive load measures did not find any significant differences between groups. First, the lack of findings may have been due to reduced statistical power caused by the PT activity scoring issue reducing the sample size.

Second, measuring the three types of cognitive load has been a major challenge for the field over the last 20 years (Sweller, 1988; van Merriënboer & Sweller, 2005). This study used self-report measures similar to those used by Mayer and Johnson (2008); however, this study was not able to administer the questions repeatedly during the study session. Administering the items after the tutorial and after the PT activity may have increased the reliability of the items. Further, the items could be administered during the tutorial and during the PT activity. While this would likely be very distracting, it might be able to provide self-report measures of changes in the subject's perceptions of the complexity of the task and the mental effort expended during the task.

Third, it is possible that the independent variable manipulations of modality and restriction did not produce differences in terms of intrinsic and germane cognitive load. According to the results in this study, additional measures other than self-report items need to be used to quantify intrinsic and germane cognitive load. However, self-report items for extraneous load were consistent with results for the transfer test. These items asked subjects to report on PT's usability. Extraneous cognitive load and usability are inherently related. Usability is the art of interface design while studying extraneous cognitive load is the science of reducing unnecessary stimuli from a visual environment. These are essentially one in the same. The field of usability has developed measures which should be integrated into educational technology research because the usability of a digital learning environment is likely to affect performance.

Lost Data

Appendix K presents the results from statistical analyses on the retention test that used the full sample (N = 146). The pretest is used as a covariate. The results must be discussed with two caveats. First, it must be noted that the reliability estimates for both these measures (pretest and retention) were below acceptable levels. Second, there was no statistical difference between retention test scores for the first and second halves of sample; however, there was a statistical difference between PT activity transfer test scores for the first and second halves of the sample. Students in the first half of sample had a different experience during the PT activity compared to the second half. The first half's PT activity scores were significantly lower than the second half's scores.

Additionally, the PT activity preceded the retention test. Thus, I did not want to use any data from the first half of sample because subjects' experiences getting low grades during the PT activity could have impacted scores on the retention test. Appendix K lists results and a short discussion of the findings from an ANCOVA (N = 146) on retention with the pretest as a covariate.

Challenges of Educational Technology Research

Numerous challenges face educational researchers, especially when using a complex, representational sim. Clark and Choi (2005) listed five principles for conducting experiments using animated pedagogical agents; however, these five principles can be used as a framework for examining the challenges experienced in this study. First, the Balanced Separation Principle states that an experiment's materials need to provide control while allowing for unique differences of the treatment to be identified. This study's materials controlled for content. The tutorial content was exactly the same for all the three modality conditions. Only the media used to deliver the content was different (text, voice, voice+text). Additionally, the PT activity was the same for both restriction conditions (restricted vs. unrestricted). Subjects had to fix a broken network. The only difference was whether parts of their interface were restricted or not. Thus, since all the materials were controlled yet different only in terms of the treatment, this study was consistent with the Balanced Separation Principle.

Second, the Variety of Outcomes Principle states that numerous measures should be used to identify and corroborate findings. This study followed previous multimedia learning research by Mayer and colleagues by using retention and transfer tests as well as measures of cognitive load. Thus, this study was consistent with the Variety of Outcomes Principle.

Third, the Robust Measurement Principle states that researchers must use measures with proven reliability and construct validity, especially when using researcher-development measures. The reliability (Cronbach's alpha) for the pretest and retention tests were both below .70. Thus, this study is not consistent with the Robust Measurement Principle giving caution to the trustworthiness of these findings. There were no significant difference for the ANCOVA (second half of sample only, N = 81) on the retention test; however, this could have been due to the low reliability on the pretest and retention measures. Future research is being planned with the first goal being to improve the reliability and construct validity of the measures used in this study.

Fourth, the Cost-Effectiveness Principle states that researchers should include the cost of developing technology into their research. Data should be collected and communicated regarding how much a technology costs to development and how the technology is worth the investment compared to not using the technology. This Return on Investment (ROI) is applicable in this study given that Cisco Networking Academy supports 700,000 users of Packet Tracer around the world. This principle is relevant to this study because Cisco must make decision on what are the most effective ways to develop tutorials that teach novices how to use PT to solve networking problems.

Fifth, the Cognitive Load Principle states that researchers must take cognitive load of the learning environment into consideration. This study examined and manipulated cognitive load and is consistent with the principle. Thus, in review of Clark and Choi's principles, this study satisfied all of the principles except the Principle of

Robust Measures. There are several reasons why the pretest and retention test had low reliability. They were researcher-developed; there could only be a limited number of items for each, and there was a limited amount of time to pilot test and refine the measures.

Recommendations for Running Educational Technology Research in a Computer Lab

Running educational technology research is very different from having subjects simply answer a paper-based survey. Data collection using educational technology is frequently complex with the potential for technical difficulties ever present. Conducting this study taught many lessons on how to plan, setup, pilot test, and run educational technology research in a computer lab within a large university setting. Planning for the study sessions should include extra time for setting up and cleaning up a lab. In addition, extra time should be allotted between pilot testing and actually running of the study for modifications. Modifications are often needed to numerous parts of the study, including the instructions, handouts, procedures, materials, and assessments. Planning how the materials will be installed on the lab's computers should also be considered.

Setting up a computer lab for educational technology research can face numerous obstacles. First, there are usually restrictions to installing software on computers in a lab; thus, administrators are needed to install all the computers in the lab with the appropriate materials, including software, content, and assessments. This generally entails contacting the lab's administrator and giving them enough time to install and test the materials. In this study, the lab's *disk image* was updated with the sim (PT) installed, two PT activity files (restricted and unrestricted), three tutorial files (text, voice, voice+text), and a link to the online survey (pretest, retention test, and self report items). A disk image is basically what is on the hard drive of each computer in the lab. Second, the sequence of steps that subjects must go through to complete the study should be simple. Files or links that are needed should be made easily accessible to subjects. For example, files should be clearly labeled and placed in a folder on the desktop of the computer. Third, the materials should be tested including all software, content, and assessments. Pilot testing is crucial to testing the most efficient ways to setup a lab. Fourth, headphones are needed when conducting modality research that uses audio for information delivery. Computers may have to be configured for headphones, and setting up headphones on 10-20 computers can take time. Thus, to increase efficiency, it is recommended to reserve a computer lab for an extended period of time, such as 6-8 hours, rather than a short period time in which headphones have to be repeatedly setup. This helps to run multiple sessions with less setup time.

Pilot testing provides a crucial opportunity to identify weaknesses in the study. Researchers should systematically check every aspect of the study. The software and content should be tested for each computer. If streaming video is used, then caution may be needed when a large sample of subjects are all downloading content at the same time. Bandwidth issues can distract subjects and even cause computer freezes. The assessments and measures need to be pilot tested as well. Reliability estimates should be calculated for each measure, and Evidence Centered Design should be used to improve the validity of assessments that are embedded within sims or games. After pilot testing has identified the issues and the modifications have been made, the study can be conducted.

Several issues can arise when running a study. Students can become confused when there is a lack of instructions. Thus, it is important to take time at the beginning of each session to clearly explain the sequence of events, directions for participation, and any Institutional Review Board information that needs to be communicated. A printed sheet of directions helps students stay on task. These recommendations are intended for conducting educational technology research in a computer lab. This study also has practical and theoretical implications which are discussed in the following section.

Implications

Practical implications

Because this study used actual materials developed and currently used by the Cisco Networking Academy, the findings may suggest instructional features that Cisco may explore to determine optimal methods for instructional delivery. Of course, prior to making any changes or investments, more research is needed to replicate and extend the findings in this study. Cisco currently uses the text-only tutorial and the unrestricted version of PT. Adding an audio narration to the tutorials has been considered by management but never implemented due to the cost of producing voice-over narrations in literally hundreds of languages. Yet, evidence from this study suggests that Cisco students may learn how to use PT better if voice is used during the tutorial rather than text. Given that the sole purpose of the “My First PT lab” tutorials is to train Cisco students how to successfully use PT, it makes sense to employ optimal methods of content delivery to train students how to use PT more effectively. If Cisco employees are more successful at using of PT, then they may perform better on real-world job tasks of building and troubleshooting computer networks. In terms of PT interface restriction, there is no evidence that restricting is better or worse than not restricting. As mentioned before, perhaps if these restricted messages were more constructive in terms of feedback and instruction, they could facilitate learning rather than be a potential source of frustration.

Theoretical implications

The finding that the voice tutorial led to higher scores on the transfer test compared to the text tutorial, within the unrestricted condition, has numerous theoretical implications. The cognitive theory of multimedia learning, and more specifically, the modality effect and the redundancy principle were supported in this study. This adds further evidence to the importance of balancing instruction across visual and verbal channels so that the user's cognitive resources are not overloaded.

Restricting access to PT's interface also has theoretical implications for cognitive load theory. It was expected that restricting access to extraneous parts of PT's interface would reduce extraneous cognitive load; however, what was not considered was the affective reaction to repeated pop-up messages with a bell that told the user the function was "Locked." This is the first study that used cognitive load to examine the effect of restricting the interface of a complex sim. Obviously, the attempt to reduce cognitive load by creating or removing restrictions from the sim's interface was not successful.

Limitations

The primary limitation in this study was losing the first half of the sample due to the scoring issue with the PT activity. This loss of data reduced power in the statistical procedures and thereby may have affected the results of this study. The reliability estimates for both the pretest and the retention test were below .7 raising issues about the integrity of the data. There were no significant findings for the pretest and the retention test; however, this could be due to the weaknesses of the measures or due to the intervention having no effect on subjects.

Second, restricting access to PT was not completely clear. Subjects in the unrestriced conditions still received “Locked” messages if they clicked on two buttons in the instructions window. Also, subjects in the unrestricted conditions could click on the “Undo” button whereas the restricted group could not. This was done at least once when a subject in the unrestricted group brought it to the researcher’s attention. This could impact the PT activity’s results by giving the unrestricted group additional scaffolding unavailable to the restricted group. There was no way to determine which subjects may have benefited in the unrestricted group. Further, for the restricted conditions, there was plenty of opportunity for getting lost within PT’s unrestricted features.

Third, this study is limited to the context of a lab setting using novices in the subject matter. Subjects spent one hour in a computer lab and were forced to participate in the study because of a subject pool class requirement. Thus, students were likely aware that there is little consequence for low performance in the subject pool study. Also, subject pool studies do not take into consideration the interacting contextual factors that impact real world, pragmatic applications of the research (Rieber, 2005). A lab setting such as this, however, provided experimental control and thus an opportunity to test the boundaries of the cognitive theory of multimedia learning and cognitive load theory. A logical sample for this study would be actual Cisco employees. This and other recommendations are included in the next section.

Recommendations for future research

Educational technology is constantly evolving; thus, on-going research is needed to build consensus on how such technology should be used to increase learning. This study examined the effects of tutorial modality and interface restriction in the context of a complex sim on retention and transfer. These findings are limited to this study; thus, additional research is needed that explores how tutorial modality and interface restriction can be optimally aligned so that learners receive information in the best way possible and so that novices have appropriate scaffolding while using a sim designed with expert functionality.

The subjects in this study were undergraduates and generally novices in the field of computer networking; however, this study could be replicated using actual Cisco Networking Academy students. The content and measures would need to assess more advanced levels of knowledge because Cisco's students use PT after first completing a few classes that establish prior knowledge. Additionally, an instructor is usually present when students first use PT. However, replicating this study with Cisco students would provide insight into the Expertise Reversal Effect (Kalyuga et al., 2003). There is evidence that experts and novices require different learning environments, and that experts understanding can actually be reduced if a learning environment is not at the appropriate level of sophistication and functionality. Collecting data that explores relationships and interactions between the level of a sim's functionality, the level of a learner's knowledge, and the optimal combination of multimedia for information delivery will inform instructional design and multimedia learning principles.

In terms of the Mayer modality line of research, a next step for this study would be to test the Mayer and Johnson (2008) revision to the redundancy principle. This would entail testing the difference between using the current voice+text condition in which the full text of narration is presented vs. a voice+text condition where the text includes only a few keywords that direct attention to critical parts of the animation. According to Mayer and Johnson, the short redundant phrases condition would outperform the full text narration condition. Yet, Mayer and Johnson used process type subject matter, such as how lighting is formed; whereas, this study first conveyed basic computer networking information then showed how to use PT. Essentially, continued research is needed to determine what are the optimal modalities to be used for different types of multimedia learning environments and for different levels of learners.

There is no existing research on the effects of restricting parts of a complex sim's interface. Thus, the research field is wide open for this topic. In particular, future studies should examine how different restriction messages affect learning. If restricting access to parts of a sim's interface is intended to scaffold the learner, then what is the optimal amount of restriction for various levels of learners? Should all extraneous functionalities be removed? Should explanatory feedback be provided? Restricting access to an interface needs further research to identify how it can be utilized to lower extraneous cognitive load, provide feedback, and improve scaffolding.

Additional future directions for this research include examining relationships between extraneous cognitive load and usability measures. Also, measures of interest and persistence along with behavioral tests of motivation are needed to shed light on how a student's affective state influences performance. A behavioral test could include having subjects use PT then take a break. During the break students can check their email, go to Facebook, continue using PT, etc. A measure of how many students continued using PT would be an indicator of motivation.

For educational technology research to develop as a field, researchers need to describe the context and specific details of their studies. There are numerous factors that affect the effectiveness of educational technology, and when a researcher does not clearly describe these factors then apples-to-apples comparisons cannot be made between studies. Perhaps more importantly, researchers should provide public access to the materials they used so that researchers around the world can use the same materials in different contexts and with different manipulations. Only by sharing and using the same sets of educational technology materials and measures can researchers begin to truly build on each others' findings. Of course, researchers will need to examine various types of materials, but the point emphasized here is that educational technology should embrace an open-source philosophy in which stakeholders around the world work together and share materials to develop software that enhances learning.

Conclusion

Ultimately, the conclusions that can be made from this study are limited due to the challenges that occurred during data collection, especially the low reliability of the pretest and retention test and the PT activity scoring issue. The findings are also limited to the study's context, e.g., undergraduates in a lab. However, this study explored a relevant topic and used an innovative methodology to examine relationships between cognitive load and complex sims. At the heart of this study is how to introduce a complex sim so that students will not only learn but also persist. The research informs two modern educational technology questions. First, what is the best instructional delivery method, in terms of text, voice, and animation or screencasts, for training novice students how to use a complex sim? Second, what effect does restricting access to certain parts of a sim's interface have on students' performance accomplishing problem solving tasks. Implications from this research affect instructional technology designers and developers of complex sims. Programmers generally do not produce novice and experts versions due to costs of programming; however, based on the cognitive load theory, novices would experience less extraneous cog load if unneeded interface elements were removed (Kalyuga et al., 2003; Mayer, 2005; Sweller & Chandler, 1994).

The terms *21st Century learner* and *21st Century learning environment* are being used pervasively today to suggest educational experiences that involve technology. PT and other complex sims are at the center of 21st Century learning experiences. These computer-based sims allow learners to practice skills and test their knowledge in ways that have no consequences. If technology is just a tool for teaching and learning, then complex sims such as PT are some of the most effective tools. Yet, these sims require training and need to be optimized through on-going research and modification cycles. If Joy (2006) and Negroponte (2007) are correct, then within 10 years children all around the world will be connected in online learning environments. It is the job of instructional technology researchers and developers to ensure that high quality, effective sims are developed so that children today and tomorrow can collaborate online to solve authentic problems.

Appendices

Appendix A – Principles of Multimedia Learning

Mayer (2005) Principles of Multimedia Learning

Groupings	Principles	Explanations
	Multimedia principle	Adding graphics (images, videos, animation) to words can improve learning
Reduce Extraneous Processing in MML	Temporal Contiguity principle	Corresponding animation and narration should be presented simultaneously rather than successively.
	Spatial-Contiguity principle	Placing text near graphics improves learning
	Signaling principle	Add cues and signal to the multimedia message that highlight the organization of essential material.
	Redundancy principle	People learn more deeply from graphics and narration than from graphics, narration, and on-screen text.
	Coherence principle	Extraneous text and graphics can hurt learning
Managing Essential Processing in MML	Segmenting principle	Present information in learner-paced segments rather than as a continuous unit.
	Pre-training principle	Prepare learners with names of concepts and basic orientation for the MML environment.
	Modality principle	Effectively utilizing the visual and auditory channels can increase working memory capacity and reduce cognitive load, i.e., Explain graphics with audio rather than text
MML Based on Social Cues	Personalization principle	Conversational tone produces deeper learner compared to formal tone.
	Voice principle	Human voice rather than a computer voice produces deeper learning
	Image principle	People do not necessarily learn more deeply when the speaker's image is on-screen rather than not on-screen.
Emerging MML principles	Animation & Interactivity principles	Learners should be able to control the pace of the animation.
	Attention-Guiding principle	Complex animations should include some sort of guide of to cue attention to critical changes in the content
	Flexibility principle	Because learners have different levels of prior knowledge, they should have choice over rather to watch the animation.
	Split-Attention principle	Avoid designs of learning environments that require the learner to split their attention between various information on-screen.

Appendix B – Pretest of computer networking knowledge

<u>Demographic info</u>

1. What is your gender?
- Male
- Female

2. What is your student classification:
- Freshman
- Sophomore
- Junior
- Senior
- Graduate student

<u>Prior Knowledge Pretest</u>

1. In terms of internet access at home, what is the difference between cable and DSL?
- DSL uses a coaxial cable, whereas cable uses a phone cable.
* DSL uses a phone cable, whereas cable uses a coaxial cable.
- DSL uses a coaxial cable, whereas cable uses a satellite.
- DSL uses a satellite, whereas cable uses a phone cable.

2. What does HTTP stand for?
- Hypertext Transfer Procedure
* Hypertext Transfer Protocol
- Hypertext Technology Procedure
- Hypertext Technology Protocol

3. Which of the following items would NOT be found on a computer's motherboard?
- RAM Memory
- BIOS chip
* DVD Burner
- CPU

4. What does WAN stand for?
- Wireless Access Network
- Wireless Area Network
- Wide Access Network
* Wide Area Network

5. What does FTP stand for?
- File Technology Procedure
- File Technology Protocol
- File Transfer Procedure
* File Transfer Protocol

6. In terms of computer security, how is a worm different than a virus?
- A worm has to attach itself to a program to infect a host.
* A worm does not have to attach itself to a program to infect a host.
- A worm is a form of virus.
- A virus is a form of worm.

7. What does DSL stand for?
* Digital Subscriber Line
- Digital Symmetric Line
- Data Subscriber Line
- Data Symmetric Line

8. Which of the following is NOT a device used for a network?
* Modulator
- Router
- Switch
- Hub

9. What does ISP stand for?
- Internet Service Procedure
- Internet Service Platform
* Internet Service Provider
- Internet Service Protocol

10. Which is the "smartest" of the network devices below?
- Ethernet cable
* Router
- Cat-5 cable
- Hub

Appendix C – Retention Test

1. Which of the following is used to test connectivity from one computer to another?
- Nslookup
- Tracert
- Netstat
* Ping

2. What type of cable connects a PC directly to a server?
- Copper non-terminating Ethernet cable
- Copper straight-through Ethernet cable
* Copper cross-over Ethernet cable
- Copper mixed-use Ethernet cable

3. Which of the following is NOT a way to learn about a device on Packet Tracer's workspace?
- Mouse over the devices to see basic configuration information about them.
- Use the Select tool to show the device configuration window
- Use the Inspect tool to view tables the network device
* Use the Setting tool to edit the device properties

4. To test a network connection, which command would you use?
- IMCP ping
* ICMP ping
- IPMC ping
- IMPC ping

5. When setting up a PC and Server for the network, which of the following is NOT essential?
- All devices need to have properly configured Fast Ethernet settings.
- All devices need to have properly configured IP address settings.
* All devices need to have properly configured ARP table settings.
- All devices need to have properly configured DNS settings.

6. Should the Port Status box be checked?
- No
* Yes
- No, except when testing the network
- Yes, except when testing the network

7. A simple connection between a PC and a Server would be considered a
- Logical Access Network
- Logical Area Network
- Local Access Network
* Local Area Network

8. What does DNS stands for?
- Domain Numerical Standard
- Domain Numerical System
- Domain Name Standard
* Domain Name System

9. How is a PDU related to a ping?
- A ping deletes a PDU between a PC and a Server.
* A ping sends a PDU between a PC and a Server.
- A PDU deletes a ping between a PC and a Server.
- A PDU sends a ping between a PC and a Server.

10. What does IP stands for?
- Information Protocol
- Information Process
* Internet Protocol
- Internet Process

11. Which of the following is not a characteristic of using a LAN?
- LANs do not need leased telecommunication lines
* LANs cost less than other types of networks
- LANs have a small geographic range
- LANs have high data-transfer rates

12. Why would an ARP table be empty?
* Because the PC and server have not been configured
- Because the PC and server have been configured
- Because the ARP table has not been configured
- Because the ARP table has been configured

13. What does PDU stands for?
- Protocol Digital User
- Protocol Digital Unit
* Protocol Data Unit
- Protocol Data User

14. When configuring a server for a network, which of the following steps does not need to be completed?
- Set the server's name
- Set the DNS server
* Set the ARP tables
- Set the IP address

Appendix D – Packet Tracer survey

1. Packet Tracer was fun.

1	2	3	4	5
Strongly Agree	Agree	Neutral	Disagree	Strongly Disagree

2. Packet Tracer's interface (visual display) was easy to use.

1	2	3	4	5
Strongly Agree	Agree	Neutral	Disagree	Strongly Disagree

3. Packet Tracer was frustrating.

1	2	3	4	5
Strongly Agree	Agree	Neutral	Disagree	Strongly Disagree

4. Packet Tracer's interface was overwhelming.

1	2	3	4	5
Strongly Agree	Agree	Neutral	Disagree	Strongly Disagree

Locked Questions for Restricted group

1. Did you receive any "Locked" messages while clicking on various parts of Packet Tracer?

- Yes
- No
- Don't know

(If Yes)

2. The "Locked" messages distracted me.

1	2	3	4	5
Strongly Agree	Agree	Neutral	Disagree	Strongly Disagree

3. Not being able to access a restricted part of Packet Tracer made me curious about why I was being restricted.

1	2	3	4	5
Strongly Agree	Agree	Neutral	Disagree	Strongly Disagree

Affective Self-Report Items

1. If you had a chance to continue exploring Packet Tracer, how eager would you be to do so?

1	2	3	4	5
Very Uneager	Uneager	Neutral	Eager	Very Eager

2. How interesting is this material?

1	2	3	4	5
Very Uninteresting	Uninteresting	Neutral	Interesting	Very Interesting

3. How much effort did you put into this study?

1	2	3	4	5
No Effort	Some Effort	Neutral	A Little Effort	All My Effort

Appendix E – Mental effort rating

This scale is intended to measure intrinsic cognitive load. These items are adapted from DeLeeuw and Mayer (2008).

1. Please rate your level of mental effort during this tutorial.

1	2	3	4	5
Extremely low mental effort	Low mental effort	Average mental effort	High mental effort	Extremely high mental effort

2. Please rate your level of mental effort using Packet Tracer.

1	2	3	4	5
Extremely low mental effort	Low mental effort	Average mental effort	High mental effort	Extremely high mental effort

3. Please rate the tutorial's level of complexity.

1	2	3	4	5
Extremely simple	Simple	Average complexity	Complex	Extremely complex

4. Please rate Packet Tracer's level of complexity.

1	2	3	4	5
Extremely simple	Simple	Average complexity	Complex	Extremely complex

Appendix F – Difficulty rating

This scale is intended to measure germane cognitive load. These items are adapted from DeLeeuw and Mayer (2008).

1. Please indicate how difficult this activity was by using the rating scale below.

1	2	3	4	5
Extremely easy	Easy	Average difficulty	Difficult	Extremely difficult

2. Please rate the how difficult the different parts of this activity were.

	1	2	3	4	5
Understanding tutorial	Extremely easy	Easy	Average difficulty	Difficult	Extremely difficult
Remembering info from tutorial	Extremely easy	Easy	Average difficulty	Difficult	Extremely difficult
Using Packet Tracer	Extremely easy	Easy	Average difficulty	Difficult	Extremely difficult

Appendix G – Task analysis for tutorial and PT activity

ECD Worksheet I: Packet Tracer Troubleshooting Activity

Title: Packet Tracer Troubleshooting Activity

Summary: Students will troubleshoot the network from the tutorial. The answer network will be missing or have incorrect elements & configurations.

Student Model Summary: Content (My First PT Lab tutorials) + Building a simple network using PT + Basic computer networking terminology & concepts.

Student Models: Troubleshooting a network, setting up a network, testing a network

Measurement Model Summary:
Scores from Retention test, Transfer test, and the two Work Products will be examined using multivariate analyses.

Focal Knowledge, Skills, & Abilities:
1. Troubleshooting the network:
   - diagnosing problems in the physical setup, and the configurations of the network.
2. Fixing the network:
   - Configuring the computers with IP addresses
   - Configuring DNS server names
3. Testing the network:
   - Sending a PDU between computers
   - Using the Scenario Window to examine the PDU

Rationale: If the tutorials are successful in teaching students how to build a single network using PT, then student should be able to troubleshoot & correct setups, sables, etc.

Tasks embedded in Fixing the Network activity:
Setting up the network
1. Turn on PC
2. Enter PC IP address
3. Enter PC DNS server
4. Click Port Status for PC
5. Turn on server
6. Enter IP address for server
7. Click Port Status for server
8. Click DNS to on for server
9. Click HTTP to on for server

Appendix H – Instructions

Instructions for Packet Tracer Study
Session: 11

ID: Ohio, Utah

Study Procedure	Time
Step 1. Pre-Survey	5 minutes
Step 2. Tutorial: *Intro to computer networking and Packet Tracer*	10 minutes
Step 3. Packet Tracer activity: *Fix a network* (Save file…)	10 minutes
Step 4. Post-Survey	15 minutes

*** Do not log out
*** Do not close any windows or programs (Packet Tracer)
*** Raise your hand for help

Instructions
Step 1. Pre-Survey of computer networking knowledge
Go to your desktop. Open Packet Tracer folder. Double click on the icon that says "Step 1". Complete this survey. You will answer 16 questions then come to a page saying "STOP". Do NOT close the survey. You will return to this browser window after the Packet Tracer activity.

Step 2. Tutorial: *Intro to computer networking and Packet Tracer*
In the same folder, double click on file labeled Step 2. A browser window will open.
Select: **Step 2-Ohio**
Watch the 10 minute video. Do not take notes.

Step 3. Packet Tracer activity: *Fix a network*
In the same folder, double click on the file labeled: **Step 3-Utah**

You are going to open a file that has a Packet Tracer network that needs to be configured so that it works. Your job is to configure it using the information below. Hint: There are 9 things to do.

You have 10 minutes. At the end of 10 minutes, SAVE your work:

Save to Desktop As: Session # - States – Your PC's ID# (top-front corner of PC)
Example: "11-Ohio Utah-12"

Network Information

Client	Server
IP address: 192.168.0.110	IP address: 192.168.0.105
DNS server: 192.168.0.105	DNS server: 192.168.0.105
	Enable HTTP

Step 4. Post-Survey
Return to the online survey. Click "Next" to begin the post-survey.

** Make sure you sign in and get your certificate of completion **

Appendix I – Consent form

Title: Evidence Centered Design for Interactive Environments: Assessment of Learning with Packet Tracer

IRB Protocol #:

Role in Study	Name	Job Title / Department	Telephone	Email
Primary Investigator	Michael Mayrath	Graduate Student, Ed Psych	789-7363	mmayrath@mail.utexas.edu
Co-Investigator	Priya Nihalani	Graduate Student, Ed Psych	789-4526	p_k_nihalani@yahoo.com
Faculty Sponsor	Dan Robinson	Professor, Ed Psych	471-2748	dan.robinson@mail.utexas.edu

You are being asked to participate in a research study. 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 any questions you may have about the study. Please read the information below and ask any questions before deciding whether or not to take part. If you choose not to complete this study, an alternate assignment will be provided by the subject pool coordinator. The alternate assignment results in equal credit as completing this study.

Your participation is entirely voluntary. You can refuse to participate without penalty or loss of benefits to which you are otherwise entitled. You can stop your participation at any time and your refusal will not impact current or future relationships with UT Austin or participating sites. To do so simply tell the researcher you wish to stop participation. The researcher will provide you with a copy of this consent for your records.

The purpose of this study is to examine the benefits and effectiveness of various uses of technology in educational settings. As part of the project, you will participate in a pre-assessment survey about technology, learn about topics via different forms of technology in one, 2-hour technology lab session, and participate in follow up surveys after the session.

If you agree to this study, we will ask you to do the following things:
- Complete the pre-assessment survey (approximately 15 minutes)
- Participate in the technology in one, one-hour technology lab. session, (60 minutes)
- Complete Survey Instrument 2 near the end of the lab session (approximately 30 minutes)

We are also asking for your permission to collect the following:
- Permit us to collect the following from your instructor: current major, gender, age

NOTE: Any identifying information will be removed from all data once it is collected and the data will be kept confidential and private. This data will inform our research but giving us permission to use it is not a mandatory part of the study. Regardless of your decision to give permission for us to use the data, you can still earn full credit for participating in our study.

Total estimated time to participate in study is: No more than 2 hours.

Risks of being in this study:
- Educational lessons using technology may pose some risks to participants. These may include: difficulty concentrating, some unintentional stress due to the controlled nature of the technology

laboratory setting, discomfort of sitting with earphones or headsets on for an extended period of time; possible headaches due to extended viewing or listening times on the technical equipment.

Benefits of being in this study:

- Possibility of gaining a better understanding of personal preferences as related to technology use in the classroom or beyond.
- Increased comfort when communicating / connecting with instructors or peers when using a variety of technology based instructional methods.
- May learn something about how technology works.

Compensation:

- Participants receive credit through participation in subject pool.
- Non-subject pool participants receive additional points towards class credit.

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 our 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 (IRB), 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 required 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.

Contacts and Questions:

If you have any questions about the study please ask now. If you have questions later, want additional information, or wish to withdraw our participation call the researchers conducting the study. Their names, phone number and email addresses are at the top of this document. If you have questions about your rights as a research participant, complaints, concerns, or questions about the research please contact Jody Jensen, Ph.D., Chair, The University of Texas at Austin Institutional Review Board for the Protection of Human Subjects at (512) 232-2685 or the Office of Research Support and Compliance at (512) 471-8871 or email orsc@uts.cc.utexas.edu.

You will be given a copy of this information for keep for your records.

Statement of Consent:

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.

Signature:____ Date:_____________

____ Date:_____________
Signature of Person Obtaining Consent

____ Date:_____________
Signature of Investigator

Appendix J – Certificate of completion

Subject Pool
Certificate of Completion

March ________, 2009

This is to verify that ________________________________ has completed 3 hours (number of hours of your study) of his/her Subject Pool Requirement in the **spring semester of 2009.** If there are any questions or problems, please contact me at mmayrath@mail.utexas.edu.

Thank you,

Michael Mayrath

Factors that Affect Learning in Complex Simulation Environments

*** **ATTENTION:** Students, please this form in a safe place. This statement demonstrates proof that you have participated in the above study. In case of a mix-up, it is your means of providing documentation that you have participated in the above study.

Appendix K - Results from full sample

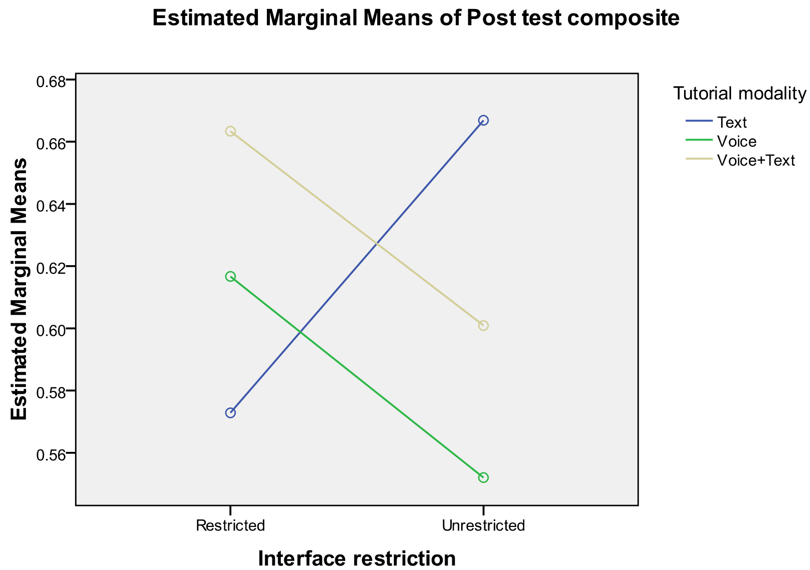
Retention Test - Significant two-way interaction

A significant interaction was found between modality and restriction on retention, $F(1, 139) = 4.199$, $p < .05$, Cohen's F = .24. The two voice conditions scored better on the retention test when they used the restricted Packet Tracer. Restricting PT's interface was good for voice-only and voice+text but not for text-only. In contrast, Leaving PT's interface unrestricted was good for text-only but bad for voice-only and voice+text. Last, voice+text scored better than voice-only. Figures 1 and 2 illustrate the modality (text, voice, voice+text) by restriction (restricted vs. unrestricted) interaction, and Table 1 below lists the results of the 2 x 3 ANCOVA on the retention test for the full sample of 146.

Table 1 – ANCOVA Source Table: Retention test interaction (N = 146, both 1st and 2nd halves).

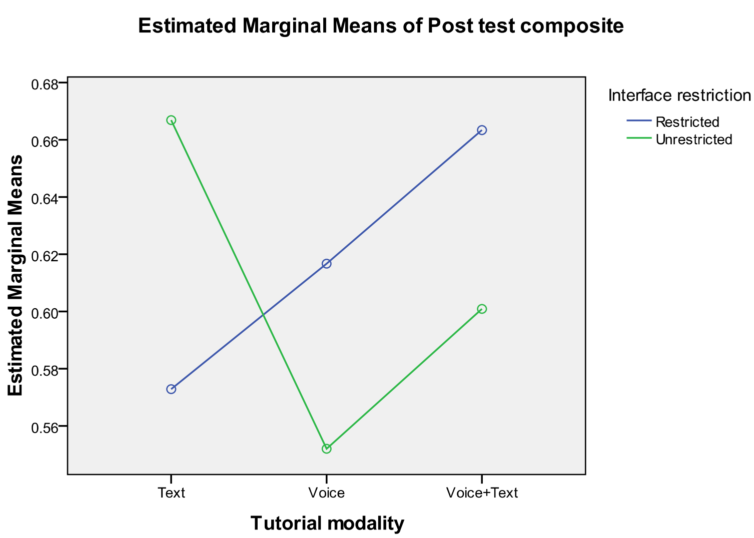
Source	Sum of Squares	df	Mean Square	F	Sig.	Cohen's f^2
Restrict	.004	1	.004	.187	.666	.04
Modality	.059	2	.030	1.240	.292	.13
Interaction	.200	2	.100	4.199	.017	.24
Error	3.312	139	.024	-----		

Figure 1 – Interaction between Modality and Restriction on Retention test.



Covariates appearing in the model are evaluated at the following values: Pre test composite = .5171

Figure 2 – Interaction between Restriction and Modality on Retention test.



Covariates appearing in the model are evaluated at the following values: Pre test composite = .5171

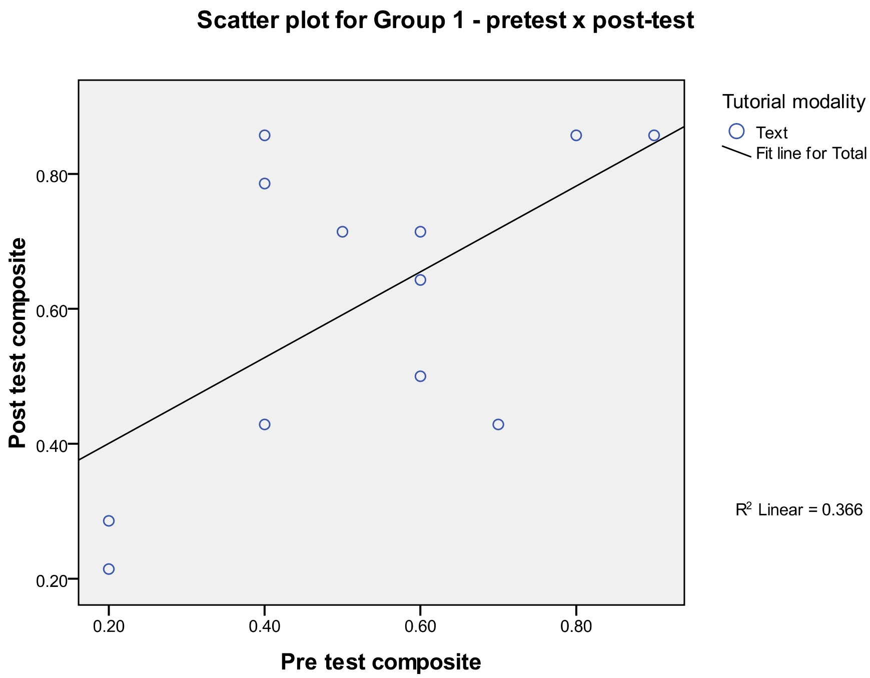
Univariate analysis

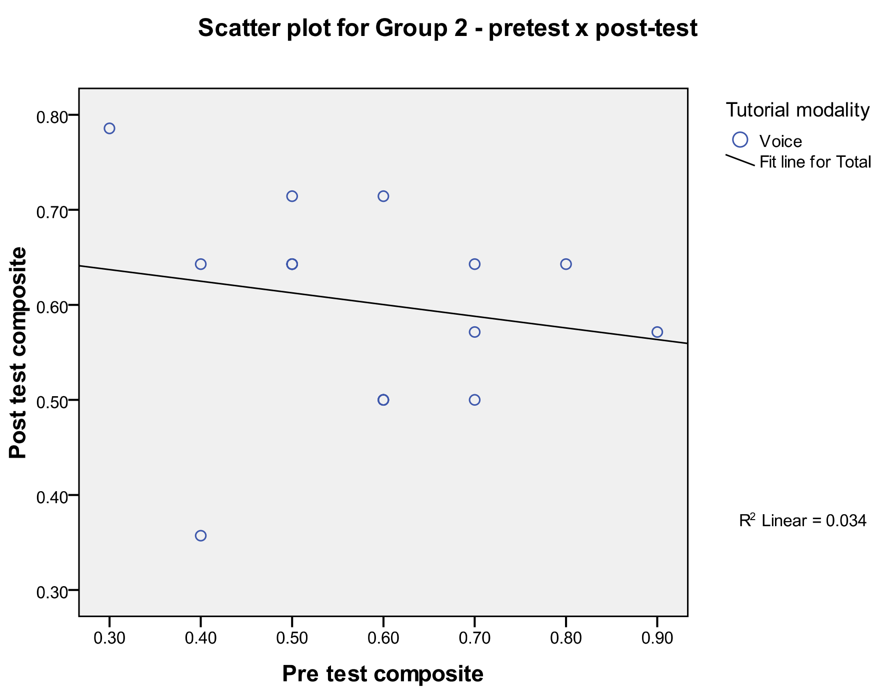
There were no main effects found for restriction on the retention test, $F(1, 139) = .187$, $p > .05$, Cohen's F = .04. The mean for restricted ($M = .618$a, $SD = .16$) was no significantly different from the mean for unrestricted ($M = .607$a, $SD = .17$). In addition, no main effects were found for modality, $F(1, 139) = 1.240$, $p > .292$, Cohen's F = .13. Voice+Text scored highest on the retention test ($M = .632$a, $SD = .17$) followed by Text-only ($M = .621$a, $SD = 17$) and lowest was Voice-only ($M = .586$a, $SD = 15$). The two text conditions outperformed the voice-only condition, and voice+text outperformed text-only. Table 2 below lists adjusted means for modality and restriction groups.

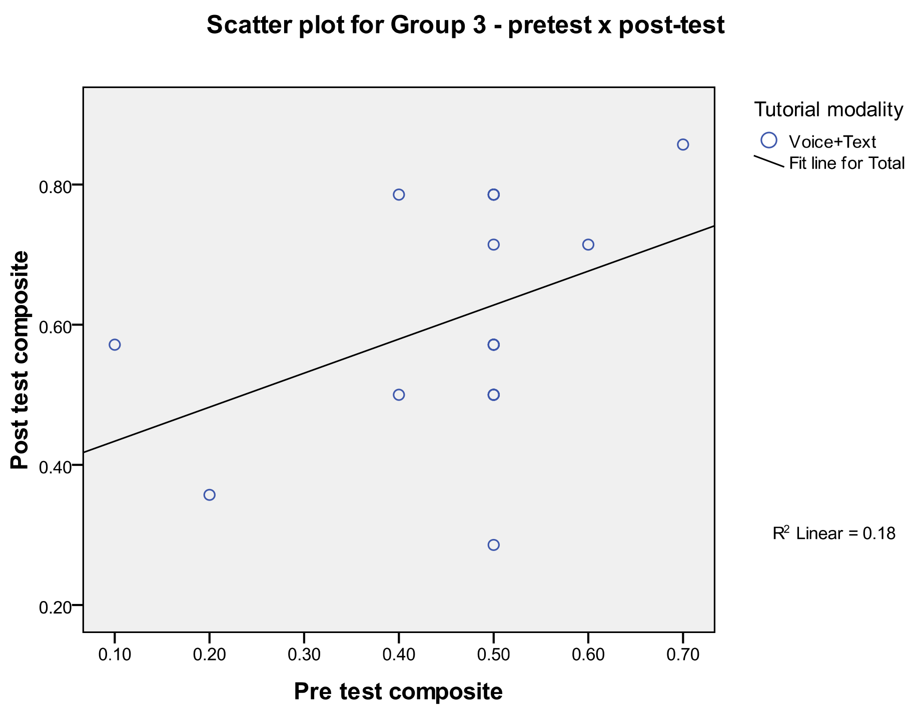
Table 2 – Means and Standard deviations for Modality and Restriction on Retention

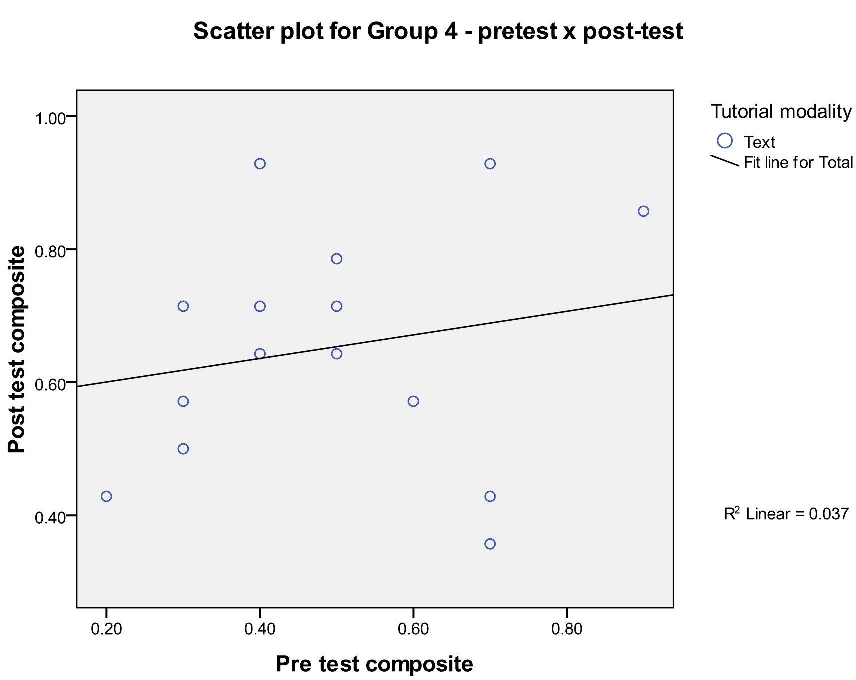
Adjusted Means, Standard Deviations, and Cell sizes				
Tutorial modality	**Interface restriction**	**M**	**SD**	**N**
Text	Restricted	.573a	.19	24
	Unrestricted	.667a	.15	25
	Total	.621a	.17	49
Voice	Restricted	.617a	.12	25
	Unrestricted	.552a	.17	24
	Total	.586a	.15	49
Voice+Text	Restricted	.663a	.16	24
	Unrestricted	.601a	.18	24
	Total	.632a	.17	48
Total	Restricted	.618a	.16	73
	Unrestricted	.607a	.17	73
	Total	.613a	.17	146

Appendix L – Scatter plots for pretest and retention by group

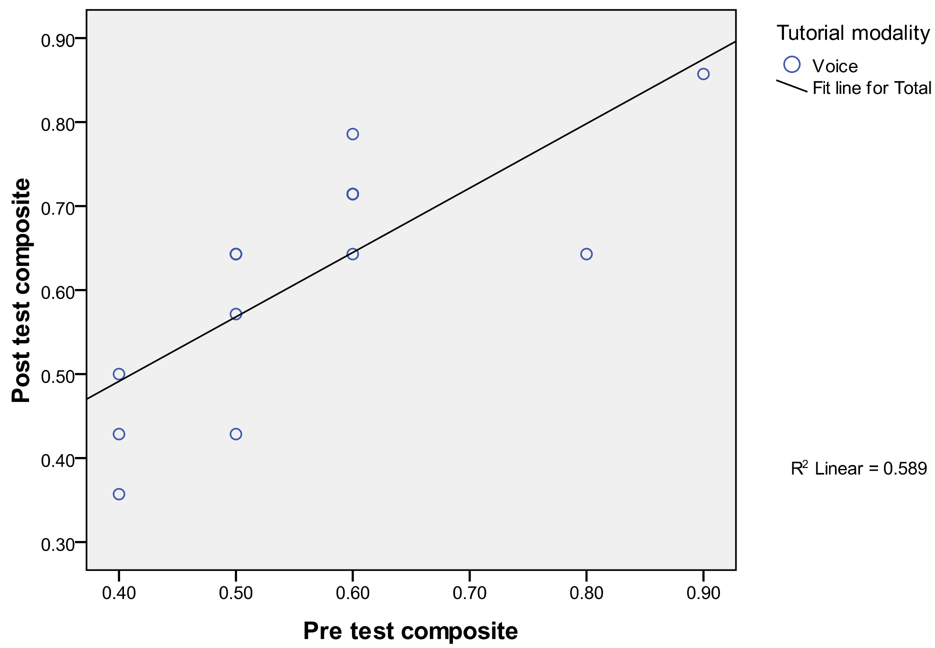




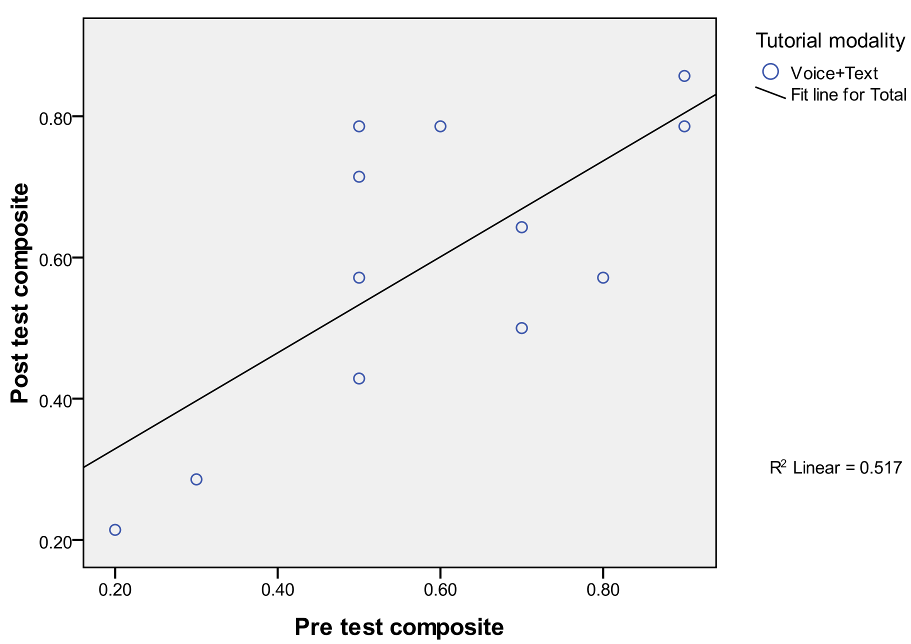

Scatter plot for Group 3 - pretest x post-test
Tutorial modality
Voice+Text
Fit line for Total
R2 Linear = 0.18
Post test composite
Pre test composite
0.80
0.60
0.40
0.20
0.10
0.20
0.30
0.40
0.50
0.60
0.70


Scatter plot for Group 4 - pretest x post-test
Tutorial modality
Text
Fit line for Total
R2 Linear = 0.037
Post test composite
Pre test composite
1.00
0.80
0.60
0.40
0.20
0.40
0.60
0.80

Scatter plot for Group 5 - pretest x post-test



Scatter plot for Group 6 - pretest x post-test



References

Atkinson, R., & Wilson, H. (1968). Computer-Assisted Instruction. *Science, 162*(3849), 73-77.

Baddeley, A. D. (1986). Working memory. New York, Oxford University Press.

Behrens, J. T., Mislevy, R. J., Bauer, M., Williamson, D. M., & Levy, R. (2004). Introduction to evidence centered design and lessons learned from its application in a global e-learning program. *The International Journal of Testing,* 4, 295-301.

Bloom, B. (1956). Taxonomy of educational objectives: cognitive domain: v. 1.

Bransford, J. (2003). *How People Learn: Brain, Mind, Experience, and School*: National Academy Press.

Brunken, R., Plass, J., & Leutner, D. (2003). Direct Measurement of Cognitive Load in Multimedia Learning. *Educational Psychologist, 38*(1), 53-61.

Cheon, J., & Grant, M. (2008). A cognitive load approach to metaphorical interface design: Reconsidering theoretical frameworks. *Technology and Teacher Education Annual,19*(2),1054-1059.

Clark, R., & Choi, S. (2005). Five Design Principles for Experiments on the Effects of Animated Pedagogical Agents. *Journal of Educational Computing Research, 32*(3), 209-225.

Clarke, J. & Dede, C. (October, 23, 2007). Data Mining as an Emerging Means of Assessing Student Learning. *EduCause Annual Conference*. Retrieved February 13th 2009 from http://connect.educause.edu/Library/Abstract/DataMiningasanEmergingMea/45355?time=1234512761

Cohen, J. (1977). *Statistical power analyses for the behavioral sciences*. New York: Academic Press.

Cuban, L. (2001). *Oversold and underused: computers in the classroom*: Harvard University Press Cambridge.

Dede, C. (1995). The Evolution of Constructivist Learning Environments: Immersion in Distributed, Virtual Worlds. *Educational Technology*. Saddle Brook, NJ. 35-46.

Dede, C., Clarke, J., Ketelhut, D., Nelson, B., & Bowman, C. (2005). Students'motivation and learning of science in a multi-user virtual environment. Paper presented at the *American Educational Research Association.* Montreal, Canada.

DeLeeuw, K., & Mayer, R. (2008). A Comparison of Three Measures of Cognitive Load: Evidence for Separable Measures of Intrinsic, Extraneous, and Germane Load. *Journal of Educational Psychology, 100*(1), 223-234.

Dieterle, E., & Clarke, J. (2007). Multi-user virtual environments for teaching and learning. In M. Pagani (Ed.), *Encyclopedia of multimedia technology and networking* (2nd ed). Hershey, PA: Idea Group, Inc. Retrieved February 29th, 2009 from http://muve.gse.harvard.edu/rivercityproject/documents/MUVE-for-TandL-Dieterle-Clarke.pdf

Ertmer, P. (2005). Teacher pedagogical beliefs: The final frontier in our quest for technology integration? *Educational Technology Research and Development, 53*(4), 25-39.

Frezzo, D., & Stanley, K. (2005). Knowledge Representations driving the Design of Computerized Performance Assessments in a complex simulated environment. Paper presented at the *American Educational Research Association*. Retrieved February 29th, 2009 from
http://www.education.umd.edu/EDMS/mislevy/CiscoAERA2005/AERA_KR_FrezzoStanley.pdf

Gee, J. (2003). *What video games have to teach us about learning*: New York: Palgrave.

Ginns, P. (2005). Meta-analysis of the modality effect. *Learning And Instruction, 15*(4), 313- 331.

Gelsinger, P. (2008). From Peta FLOPS to Milli Watts. Presentation at *Intel Developer Forum*.Retrieved February 13th 2009 from
http://download.intel.com/pressroom/kits/events/idfspr_2008/2008_0402_IDF-PRC_Gelsinger_EN.pdf

Gredler, M. (1996). Educational games and simulations: A technology in search of a (research) paradigm. *Handbook of research for educational communications and technology*, 1, 521-540.

Gredler, M. (2004). Games and simulations and their relationships to learning. *Handbook of research on educational communications and technology*, 571-581.

Harskamp, E. G., Mayer, R. E., & Suhre, C. (2007). Does the modality principle for multimedia learning apply to science classrooms? *Learning and Instruction, 17*(5), 465-477.

Herrington, J. (2007). Authentic Learning Supported by Technology: Ten suggestions and cases of integration in classrooms. *Educational Media International, 44*(3), 219-236.

Herrington, J. & Oliver, R. (1999). Using Situated Learning and Multimedia to Investigate Higher-Order Thinking. *Journal of Interactive Learning Research. 10*(1), 3-24.

Jonassen, D., & Rohrer-Murphy, L. (1999). Activity theory as a framework for designing constructivist learning environments*. Educational Technology Research and Development, 47*(1), 61-79.

Joy, B. (2006). What I'm worried about, what I'm excited about. *Technology, Entertainment, and Design conference*. Long Beach, CA. Retrieved February 12th 2009 from
http://www.ted.com/index.php/talks/bill_joy_muses_on_what_s_next.html

Kalyuga, S., Ayres, P., Chandler, P., & Sweller, J. (2003). The Expertise Reversal Effect. *Educational Psychologist, 38*(1), 23-31.

Lee, H.L., Plass, J.L., and Homer, B.D. (2006). Optimizing cognitive load for learning from computer-based science simulations. *Journal of Educational Psychology 98*(4): 902-913.

Lave, J., & Wenger, E. (1991). *Situated Learning: Legitimate Peripheral Participation*: Cambridge University Press.

Lisle, A. (2007). Assessing learning styles of adults with intellectual difficulties. *Journal of Intellectual Disabilities*, *11*(1), 23-45.

Mackey, T., & Ho, J. (2008). Exploring the Relationships between Web Usability and Students. *Computers & Education, 50*(1), 386-409.

Mayer, R. (2005). *The Cambridge Handbook of Multimedia Learning*: Cambridge University Press.
Mayer, R., & Johnson, C. (2008). Revising the Redundancy Principle in Multimedia Learning. *Journal of Educational Psychology, 100*(2), 380-386.
Mayer, R. E. (2001). *Multimedia learning*. New York: Cambridge University Press.
Mayer, R. E. (2002). Cognitive Theory and the Design of Multimedia Instruction: An Example of the Two-Way Street Between Cognition and Instruction. *New Directions for Teaching and Learning, 2002*(89), 55-71.
Meshkati, N., & Loewenthal, A. (1988). An eclectic and critical review of four primary mental workload assessment methods: A guide for developing a comprehensive model. *HumanMental Workload*, 251-268.
Miller, G. A. (1956). The magical number seven, plus or minus two: Some limits on our capacity for processing information. *Psychological Review*, 63, 81-97.
Mislevy, R. J. (2006). Cognitive psychology and educational assessment. In R. L. Brennan (Ed.), *Educational measurement* (4th ed., pp. 257-305). Westport, CT: American Council on Education/Praeger Publishers.
Mislevy, R. J., Steinberg, L. S., & Almond, R. G. (2003). On the structure of educational assessments. *Measurement: Interdisciplinary Research and Perspectives*, 1, 3-67.
Moore, G. E. (February 10, 2003). No Exponential is Forever...but We Can Delay 'Forever'. Presentation at *International Solid State Circuits Conference* (ISSCC). Retrieved February 13th 2009 from http://www.intel.com/cd/corporate/techtrends/emea/eng/209729.htm
Moreno, R. (2006). Does the modality principle hold for different media? A test of the method-affects-learning hypothesis. *Journal of Computer Assisted Learning, 22*(3), 149-158.
Moreno, R., & Mayer, R. (2002). Verbal Redundancy in Multimedia Learning: When Reading Helps Listening. *Journal of Educational Psychology, 94*(1), 156-163.
Moreno, R., Mayer, R., Spires, H., & Lester, J. (2001). The Case for Social Agency in Computer-Based Teaching: Do Students Learn More Deeply When They Interact With Animated Pedagogical Agents? *Cognition and Instruction, 19*(2), 177-213.
Negroponte, N. (2007). One Laptop per Child, two years on. *Technology, Entertainment, and Design conference*. Long Beach, CA. Retrieved February 12th 2009 from http://www.ted.com/index.php/talks/nicholas_negroponte_on_one_laptop_per_child_two_years_on.html
Nielsen, J. (2001). *Ten Usability Heuristics*. Retrieved February 13th, 2009 from http://katsvision.com/canm606/session_2/M2_reading2.pdf
O'Neil, H., Wainess, R., & Baker, E. (2005). Classification of learning outcomes: evidence from the computer games literature. *Curriculum Journal, 16*(4), 455-474.
Paas, F., Tuovinen, J., Tabbers, H., & Van Gerven, P. (2003). Cognitive Load Measurement as a Means to Advance Cognitive Load Theory. *Educational Psychologist, 38*(1), 63-71.
Papert, S. (1980). *Mindstorms: children, computers, and powerful ideas*: Basic Books, Inc. New York, NY, USA.

Pavio, A. (1986). *Mental representations: A dual coding approach*. Oxford, England: Oxford University Press.

Penney, C. (1989). Modality effects and the structure of short-term verbal memory. *Memory & Cognition*, 17, 398-422.

Peterson, L., & Peterson, M. (1959). Short-term retention of individual verbal items. *Journal of Experimental Psychology*, 58, 193-198.

Perkins, D. (1993). Person-plus: A distributed view of thinking and learning. *Distributed cognitions: Psychological and educational considerations*, 88-110.

Perkins, D. (1995). *Smart Schools: Better Thinking and Learning for Every Child*: Free Press.

Resta, P., & Laferrière, T. (2007). Technology in Support of Collaborative Learning. *Educational Psychology Review, 19*(1), 65-83.

Rieber, L. (2005). Multimedia learning in games, simulations, and microworlds. *The Cambridge Handbook of Multimedia Learning*, 549–567.

Ross, S., & Morrison, G. (1996). Experimental research methods. In D. H. Jonassen (Ed.), Handbook of research for educational communications and technology (pp. 1148-1169). New York: Simon & Schuster Macmillan

Stevens, J. (2002). *Applied Multivariate Statistics for the Social Sciences*: Lawrence Erlbaum Associates.

Sweller, J. (1988). Cognitive load during problem solving: Effects on learning. *Cognitive Science, 12*(2), 257-285.

Sweller, J. (2005). The Redundancy Principle in Multimedia Learning. *The Cambridge Handbook of Multimedia Learning*, 159-168.

Sweller, J., & Chandler, P. (1994). Why Some Material Is Difficult to Learn. *Cognition and Instruction, 12*(3), 185-233.

Sweller, J., van Merriënboer, J., & Paas, F. (1998). Cognitive Architecture and Instructional Design. *Educational Psychology Review, 10*(3), 251-296.

U.S. Department of Education. (2008). *Harnessing Innovation to Support Student Success: Using Technology to Personalize Education*. Washington, D.C. Retrieved February 12th 2009 from
http://www.ed.gov/about/offices/list/os/technology/reports/roundtable.html

van Merriënboer, J., & Sweller, J. (2005). Cognitive Load Theory and Complex Learning: Recent Developments and Future Directions. *Educational Psychology Review, 17*(2), 147-177.

Whelan, R. (2007). Neuroimaging of cognitive load in instructional multimedia. *Educational Research Review, 2*(1), 1-12.

Zhao, Y., & Frank, K. (2003). Factors Affecting Technology Uses in Schools: An Ecological Perspective. *American Educational Research Journal, 40*(4), 807-840.

Vita

Michael Charles Mayrath attended J. J. Pearce High School, Richardson, Texas. In 1995 he entered University of Texas at Austin. He received a Bachelor of Journalism degree in May, 1999. He spent the next year traveling Australia, New Zealand, and Tahiti. Upon returning to Austin, Texas he worked in technology sales and marketing. In August, 2004, he entered the Graduate School at the University of Texas at Austin.

Permanent Address: 1112 Juniper St., Austin, Texas, 78702

This manuscript was typed by Michael C. Mayrath.