THE ACQUISITION OF PROCEDURAL SKILLS:
AN ANALYSIS OF THE
WORKED EXAMPLE EFFECT
USING ANIMATED DEMONSTRATIONS

by

David Lewis

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

Major Professor: Ann Barron, Ed.D.
Michael Coovert, Ph.D.
William Kealy, Ph.D.
Jeffrey Kromrey, Ph.D.

June 24, 2007

Keywords: Cognitive Load, Learning Efficiency, Animated Demonstration

© Copyright 2007, David Lewis
# TABLE OF CONTENTS

<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>TABLE OF CONTENTS</td>
<td>i</td>
</tr>
<tr>
<td>LIST OF TABLES</td>
<td>iv</td>
</tr>
<tr>
<td>LIST OF FIGURES</td>
<td>v</td>
</tr>
<tr>
<td>ABSTRACT</td>
<td>vi</td>
</tr>
<tr>
<td>CHAPTER 1 - INTRODUCTION</td>
<td>1</td>
</tr>
<tr>
<td>The significance of the problem</td>
<td>1</td>
</tr>
<tr>
<td>Purpose</td>
<td>3</td>
</tr>
<tr>
<td>Research Questions</td>
<td>3</td>
</tr>
<tr>
<td>Overview of the Methodology</td>
<td>5</td>
</tr>
<tr>
<td>Limitations</td>
<td>6</td>
</tr>
<tr>
<td>Delimitations</td>
<td>7</td>
</tr>
<tr>
<td>Terminology</td>
<td>8</td>
</tr>
<tr>
<td>Summary</td>
<td>11</td>
</tr>
<tr>
<td>CHAPTER 2 – LITERATURE REVIEW</td>
<td>12</td>
</tr>
<tr>
<td>Learning</td>
<td>12</td>
</tr>
<tr>
<td>Procedural and declarative learning</td>
<td>12</td>
</tr>
<tr>
<td><strong>Procedural knowledge acquisition</strong></td>
<td>13</td>
</tr>
<tr>
<td>Expertise and schema theory</td>
<td>15</td>
</tr>
<tr>
<td><strong>Information processing and human problem solving</strong></td>
<td>15</td>
</tr>
<tr>
<td>Chess and human expertise</td>
<td>16</td>
</tr>
<tr>
<td>Schema theory</td>
<td>18</td>
</tr>
<tr>
<td>Schemas as problem categories</td>
<td>18</td>
</tr>
<tr>
<td>Cognitive load theory</td>
<td>19</td>
</tr>
<tr>
<td>Element interactivity</td>
<td>21</td>
</tr>
<tr>
<td>Managing cognitive load</td>
<td>22</td>
</tr>
<tr>
<td>Problem solving and cognitive load</td>
<td>23</td>
</tr>
<tr>
<td><strong>Means-ends analysis and cognitive load</strong></td>
<td>23</td>
</tr>
<tr>
<td>Worked examples</td>
<td>25</td>
</tr>
<tr>
<td>When are worked examples warranted?</td>
<td>26</td>
</tr>
<tr>
<td>The design of effective worked examples</td>
<td>27</td>
</tr>
<tr>
<td><strong>Problem completion</strong></td>
<td>28</td>
</tr>
<tr>
<td>The variability of worked examples</td>
<td>29</td>
</tr>
<tr>
<td>Split-attention</td>
<td>30</td>
</tr>
<tr>
<td>Cognitive load and multimodal instruction</td>
<td>32</td>
</tr>
<tr>
<td>The mode of instruction</td>
<td>33</td>
</tr>
<tr>
<td>Animation as an instructional strategy</td>
<td>34</td>
</tr>
<tr>
<td>Chapter 3 Methods</td>
<td>Page</td>
</tr>
<tr>
<td>-------------------</td>
<td>------</td>
</tr>
<tr>
<td>Research Design</td>
<td>68</td>
</tr>
<tr>
<td>Sample</td>
<td>68</td>
</tr>
<tr>
<td>Research questions</td>
<td>69</td>
</tr>
<tr>
<td>Expectations</td>
<td>69</td>
</tr>
<tr>
<td>Materials</td>
<td>70</td>
</tr>
<tr>
<td>Introductory survey</td>
<td>70</td>
</tr>
<tr>
<td>Overview</td>
<td>71</td>
</tr>
<tr>
<td>Instructional Materials</td>
<td>71</td>
</tr>
<tr>
<td>The Picnic Problem</td>
<td>73</td>
</tr>
<tr>
<td>Procedure</td>
<td>74</td>
</tr>
<tr>
<td>Data collection</td>
<td>75</td>
</tr>
<tr>
<td>The learning environment</td>
<td>75</td>
</tr>
<tr>
<td>Acquisition phase</td>
<td>75</td>
</tr>
<tr>
<td>Performance Phase</td>
<td>77</td>
</tr>
<tr>
<td>Dependent variables</td>
<td>78</td>
</tr>
<tr>
<td>Performance Time</td>
<td>78</td>
</tr>
<tr>
<td>Accuracy</td>
<td>78</td>
</tr>
<tr>
<td>Relative condition and learning efficiency</td>
<td>79</td>
</tr>
</tbody>
</table>
LIST OF TABLES

Table 1. Terminology.......................................................................................................... 8

Table 2. Methods for measuring cognitive load .............................................................. 58

Table 3. Relative instructional efficiency and learning efficiency ................................. 62

Table 4. Research questions one and two described as hypotheses............................... 70

Table 5. Instructional materials......................................................................................... 72

Table 6. Week one- the Mr. Potatohead problem............................................................ 73

Table 7. Week two - the picnic problem............................................................................ 74

Table 8. Relative condition efficiency and learning efficiency ........................................ 82
LIST OF FIGURES

Figure 1. Cognitive load over time (adapted from Paas et al, 2003) ......................... 20

Figure 2. Split-Attention Diagrams ........................................................................... 31

Figure 3. Performance Time (adapted from Waterson and O’Malley, 1993) .......... 41

Figure 4. Structured and ill-structured domains (adapted from Jonassen et al, 1993) 52

Figure 5. Means of test performance (adapted from Tuovinen & Sweller, 1999) .... 56

Figure 6. Relative condition efficiency (E) ................................................................. 60

Figure 7. Relative condition efficiency (adapted from Paas and van Merriënboer, 1993) 60

Figure 8. Overall or 3D efficiency .............................................................................. 64

Figure 9. Training efficiency ..................................................................................... 64

Figure 10. Performance efficiency ............................................................................ 65

Figure 11. Paas and van Merriënboer’s efficiency formula ....................................... 83
ABSTRACT

Most educators suggest active rather than passive instruction. In contrast to this conventional wisdom, Sweller and Cooper found that those learners that passively studied worked examples were significantly more efficient than learners who actively solved problems (Sweller & Cooper, 1985; Cooper & Sweller, 1987). This study tests this “worked example effect” (and the associated “variability effect”) using animated demonstrations. This study considers the learner performance of groups that either: study animated demonstrations, practice procedures, or study demonstrations and practice procedures. The study uses a MANOVA to compare learner performance time and accuracy given the above groups, one week after initial instruction. In addition this study considers instructional condition efficiency and learning efficiency.
CHAPTER 1 - INTRODUCTION

Educators have traditionally questioned a passive reception of instruction and instead suggest an active construction of knowledge (Bruner, 1961, Dewey, 1916/1997; Jonassen, 1991; Wittrock, 1974). Even though this philosophy has a rich literature, there is a wealth of empirical evidence to suggest otherwise. Sweller and Cooper were perhaps the first to find that if learners passively studied worked examples during early schema acquisition, that they would significantly out-performed their peers, who had learned the same procedures through active problem solving (Sweller & Cooper, 1985, Cooper & Sweller, 1987). This has since been described as the “worked example effect” (Sweller, 1993; Sweller, van Merriënboer, & Paas, 1998). This effect has been replicated by many researchers in a variety of circumstances (Sweller, 1988, Carroll, 1994; Paas, 1992; Paas & van Merriënboer, 1994; Quilici & Mayer, 1996; Zhu & Simon, 1987).

The present study considers the performance of those that study animated demonstrations, a form of animated worked example (Lewis, 2005). Sweller and Cooper’s studies demonstrate that learner performance is positively affected during worked example-based instruction. Specifically, Sweller recommends that during the early stages of schema acquisition, that the actual performance of a procedure may be detrimental to learning, because it adds extraneous load to an already complex learning environment (Sweller, 1993).

The significance of the problem

Perhaps the most important reason to study animated demonstrations is this form of instruction is generalizable to all computer applications. Instructional designers can
use recording software to record any computer-based procedure. In addition, animated demonstrations may be designed so that they can make efficient use of both visual and verbal modalities [dual coding theory (Paivio, 1971, 1978)] allowing for multimedia learning (Mayer, 2001). As we all know computer-based procedures have become very important in our knowledge worker society, so an efficient method of instruction that is generalizable to all computer-based procedures is very valuable indeed.

Thus it is no surprise that animated demonstrations have become increasingly common, and used by both educators and industry. Well known companies like Bank of America and Amazon.com, are using animated demonstrations as a way to teach clients how to use their online services. Microsoft, perhaps the world’s most famous software company, has begun to incorporate “demos” (animated demonstrations) into its Microsoft Office® products, as training and support. They also offer this training through a website called the “Office Demo Showcase” (http://office.microsoft.com/en-us/FX010804491033.aspx). Finally, training groups like “Element K” are also beginning to offer animated demonstrations as an “online training service.

Lewis (2005) proposed animated demonstrations are a form of animated worked examples. Even though Sweller and his associates have found worked examples to be an efficient and effective form of instruction (Sweller & Cooper, 1985, Cooper & Sweller, 1987), few (if any) Cognitive load researchers have compared the performance of those studying animated demonstrations to those using discovery-based problem solving. Thus it has yet to be determined empirically if animated demonstrations will exhibit the worked example effect; and thus truly are animated worked examples.
Purpose

Given the prevalence of animated demonstrations, it is imperative that the efficiency and effectiveness of this presentation form be tested empirically. Thus the purpose of this dissertation project is to test the “worked example effect” [and “the variability effect” (an extension of the worked example effect)] using animated demonstrations.

Sweller and his associates have repeatedly found evidence that instructional materials that demonstrate a procedure are more efficient than discovery-based problem solving. Sweller (1988) suggests this is because during early schema acquisition, learners who interact with content are engaging in problem-solving search, a non-schema forming activity. Thus Sweller and Cooper developed the worked example strategy in order to prevent problem solving search. This study suggests that animated demonstrations also eliminate problem solving search, thus lowering the extraneous cognitive load associated with problem solving, promoting learner performance. But this claim must be tested.

Finally, while worked examples are an efficient means of introducing procedure-based learning, but the long-term retention of this type of learning, may not be as durable, when compared with learning given exploration or discovery-based problems solving (Tuovinen & Sweller, 1999). Thus this must also be tested.

Research Questions

In this study it is proposed that learners, who passively study animated demonstrations during early schema acquisition, will outperform those who actively solve problems. Thus this dissertation questions the active learning assumption; for it expects that the “worked example effect” holds for animated demonstrations. In addition,
this project questions the use of discovery-practice following animated demonstrations. It is expected that those learners who engage in problem solving immediately following instruction with animated demonstrations, may also have a performance decrement given a delayed retention test.

This dissertation compares four instructional conditions: animated demonstration only (demo only), demonstration plus practice (demo+practice), a second demonstration plus practice (demo2+practice), and finally a practice only condition (practice only). The following research questions are based on the possible outcomes of instruction using animated demonstrations versus those using discovery problem solving:

_**Question 1:** Will learners using one of the instructional conditions complete tasks in significantly less time than learners in the other conditions? (Performance time)_

_**Question 2:** Will learners using one of the instructional conditions be significantly more accurate? (Accuracy)_

_**Question 3:** Will one of the instructional conditions be significantly more efficient during initial training? (Learning efficiency)_

_**Question 4:** Will one of the instructional conditions be significantly more efficient during delayed test? (Instructional efficiency)_

_**Question 5:** Will the group performance from one of the instructional conditions be significantly more efficient during delayed test? (Performance efficiency)_

It is expected that a comparison of the demo only and practice only conditions will answer questions concerning the worked example effect. A comparison of the demo only and demo+practice groups, will answer the question concerning practice following
an animated demonstration. Finally a comparison of conditions demo+practice and demo2+practice groups, will answer questions concerning the variability effect.

Overview of the Methodology

As the literature review for this study explains, many researchers studying learning with animation typically gather data through pencil and paper tests, and thus usually only assess conceptual or declarative knowledge. But procedure-based learning is most accurately assessed through observation. Observation is the fundamental basis of science, but is often the most under used and under valued method of data collection given human performance (Pershing, Warren, & Rowe, 2006).

This project suggests educational research may be extended to studying procedure-based learning with animated demonstrations, given observation is used as a means of data collection. To accomplish this goal, this study intends to utilize the screen capture technologies of Human Computer Interaction (HCI) research, to monitor learner behavior over time, in order to assess procedural learning.

TechSmith Morae, a usability software is the primary tool for data collection (Techsmith, 2004) (note Appendices A, B & C). This usability software is composed of two components, Morae Recorder and Manager. Morae Recorder acts like a video camera, to record learner interaction with a computer, to produce a coded movie file. Certainly this movie file is a visual record of a learner’s onscreen actions, but, in addition, Morae Recorder encodes a database of all user actions (mouse clicks, keyboard entries, & window events) into the coded movie file.
Morae Recorder is installed on a lab computer and turned on before a learner sits down to interact with the computer. This software may be completely hidden from the learner, making it a non-reactive measure (Campbell, 1957), allowing the researcher to observe learner behavior, but not be intrusive or change the nature of the behavior.

The second component of this system, Morae Manager, allows a researcher to analyze the coded movie file produced by Recorder. Researchers use Manager to analyze and document learner interaction with the computer. Typically a researcher searches the user action database for all relevant actions. In addition, researchers can then add to the database by labeling learner actions with a series of markers (small flags on the video timeline). Two markers may used together to describe a length of video (known as a segment). Each of these events, markers and segments, are automatically time stamped by the software in hundredth of seconds. This level of precision will not be necessary, as time in seconds will be sufficient for the purposes of this study.

Since all user actions are recorded in the Morae Manager video database, each learner’s file will provide all the data necessary to answer the research questions. Several outcome variables will be recorded and measured using Morae Manager (performance time, accuracy, and learner efficiency). These variables and two others will be documented within the individual’s file and then analyzed using multivariate statistics. The details of the analysis procedures are described in chapter 3.

Limitations

According to some cognitivists, learning is defined as “a change in long-term memory” (Kirschner, Sweller, & Clark, 2006, p.75). These authors made this statement because of recent advances in the cognitive sciences, specifically in brain imaging
technology. Even though this is the case, it is still difficult to directly measure changes in long-term memory without brain imaging techniques. Thus current technologies limit educational researchers to only indirectly measuring learning, by observing behavior.

This methodology (behavior analysis) has its limitations. Given Kirschner, Sweller, and Clark’s definition of learning, it will not be known if the learner has actually learned how to perform a procedure, only that they have performed the procedure. It is completely possible that learners can discover the correct series of problem solving operators and not be able to reproduce that performance. Unfortunately, even given this limitation of behavior analysis, it must be assumed that if a learner has performed a procedure, that they have learned that procedure.

The methodology of this study is fully described in Chapter 3, and may be described as “computer-supported data collection.” Certainly computer-supported data collection is perhaps the most accurate way to record user actions, but unfortunately not all user actions may be described with Morae search results. For instance, although mouse clicks (or “mousedown” events) are recorded by Morae, a “mouseup” event is not recorded. This is unfortunate because researchers must decide when the learner ends some procedures. Given this basic limitation of the recording technology, researchers must define some learner actions themselves, allowing for some measurement error.

**Delimitations**

Delimitations describe the populations to which a study’s results may be generalized (Locke, Spirduso, & Silverman, 2000). The participants in this study are pre-service teachers taking a required, sophomore level Educational technology course at a large southeastern university. This diverse group of individuals is fairly representative of
college-aged adults, although the population studied is likely to contain more females than males.

The proposed study measures learner performance given computer-based instruction. Specifically, it only measures on-screen interaction, a limited form of human-computer interaction. In addition, it only measures the behavior of novices during learning. Therefore the results of this study may only be generalized to adult learners, specifically novices, engaged in human-computer interaction.

**Terminology**

This project brings together the research of several fields of study, thus there is a broad array of terms used in this document. If necessary, please consult Table 1 for the definition of common terms.

**Table 1. Terminology**

<table>
<thead>
<tr>
<th>Term or acronym</th>
<th>Definition</th>
</tr>
</thead>
<tbody>
<tr>
<td>animated demonstration</td>
<td>A narrated animation depicting procedure-based instruction</td>
</tr>
<tr>
<td>cognitive load theory</td>
<td>John Sweller has synthesized several theories (working memory, schema acquisition, and instructional design theory) to derive his own theory of human performance given the information processing requirements of instructional materials (Sweller, 1988). Sweller and others have used this theory to predict and document a number of important learning effects associated with the complexity of instructional materials (e.g. Worked-example effect, Completion problem effect, Split-attention effect, Modality effect, Redundancy effect, Variability effect, &amp; the Element interactivity effect).</td>
</tr>
<tr>
<td>Term</td>
<td>Description</td>
</tr>
<tr>
<td>-------------------------------</td>
<td>---------------------------------------------------------------------------------------------------------------------------------------------</td>
</tr>
<tr>
<td>completion problem effect</td>
<td>Paas (1992) found that those learners who study and use partially worked-out examples (completion problems), performed significantly better (took less time with less effort), than their peers who used traditional problem solving strategies.</td>
</tr>
<tr>
<td>declarative learning</td>
<td>Declarative learning is concerned with the learning of language-based information (e.g. facts and events) (Squire &amp; Zola, 1996).</td>
</tr>
<tr>
<td>discovery learning</td>
<td>A type of learning that became popular in the 1960s. Proponents suggest that when one discovers information for oneself, he or she is more likely to remember it (Bruner, 1961).</td>
</tr>
<tr>
<td>discovery approach</td>
<td>An approach to instructional design that promotes self-guided instruction.</td>
</tr>
<tr>
<td>element interactivity</td>
<td>According to Sweller, instructional content is composed of component parts or “elements” (Sweller &amp; Chandler, 1994). Elements may be said to “interact” if there is a relationship between them, thus raising the complexity of the instruction. The total number of elements is not as important as the number of interactions between these elements.</td>
</tr>
<tr>
<td>expertise reversal effect</td>
<td>As learners become more competent, the worked-example and other cognitive load effects disappear (Kalyuga, Chandler, &amp; Sweller, 1998); this has been termed the “expertise reversal effect” (Kalyuga, Ayres, Chandler, &amp; Sweller, 2003).</td>
</tr>
<tr>
<td>expository approach</td>
<td>This is an approach to the design of instruction in which the learner is provided with the most guidance possible.</td>
</tr>
<tr>
<td>extraneous cognitive load</td>
<td>Extraneous cognitive load is that load not inherent within the activity (Chandler &amp; Sweller, 1991; Chandler &amp; Sweller, 1992), but is load that may be controlled by the instructional designer as they structure and present instructional materials (Pollock, Chandler, &amp; Sweller, 2002).</td>
</tr>
<tr>
<td>functional magnetic resonance imaging (fMRI)</td>
<td>This brain imaging technique has energized cognitive science for it allows researchers to better understand the cognitive functions of the brain. For instance, learning theorists Anderson, Albert, and Fincham (2005) have used this technique to better what areas of the brain are used during problem solving.</td>
</tr>
<tr>
<td>Term</td>
<td>Definition</td>
</tr>
<tr>
<td>-----------------------------</td>
<td>-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------</td>
</tr>
<tr>
<td>germane cognitive load</td>
<td>Germane (or Relevant) cognitive load is load directed toward schema construction (Sweller, Van Merriënboer, &amp; Paas, 1998).</td>
</tr>
<tr>
<td>HCI (human-computer interaction)</td>
<td>The reciprocal events related to the behavior of humans and computers (also known as human-computer interaction, or in instructional settings as learner interaction) (Wagner, 1994, Moore, 1989).</td>
</tr>
<tr>
<td>intrinsic cognitive load</td>
<td>Intrinsic cognitive load is the inherent level of difficulty or complexity associated with an instructional activity (Chandler &amp; Sweller, 1991; Chandler &amp; Sweller, 1992).</td>
</tr>
<tr>
<td>job aid</td>
<td>A text-based list of instructions (Rossett &amp; Gautier-Downes, 1991).</td>
</tr>
<tr>
<td>marker</td>
<td>A small flag placed on the Morae Manger timeline. It represents a researcher designated event or action. For instance it may represent the end of a user action.</td>
</tr>
<tr>
<td>modality effect</td>
<td>This effect suggests learners have superior performance given multimedia (dual modality - visual and verbal) based instructional materials (Moreno &amp; Mayer, 1999; Mousavi, Low, &amp; Sweller, 1995; Mayer, 2001; Penney, 1989).</td>
</tr>
<tr>
<td>procedural learning</td>
<td>This is skills based learning (e.g. learning how to use a computer program) (Squire &amp; Zola, 1996). When one is learning “how to” do something they are engaged in procedural learning.</td>
</tr>
<tr>
<td>problem-solving operator</td>
<td>“an action that transforms one state into another state.” (e.g. in a maze the operators are going from one location to another) (Anderson, 1993, p.36)</td>
</tr>
<tr>
<td>relative condition efficiency</td>
<td>“the observed relation between mental effort and performance in a particular condition in relation to a hypothetical baseline condition in which each unit of invested mental effort equals one unit of performance” (Paas &amp; Van Merrienboer, 1993, p. 739).</td>
</tr>
</tbody>
</table>
| schema                      | “…a structure which allows problem solvers to recognize a problem state as belonging to a particular category of problem states that normally require
particular moves.” (Sweller, 1988, p. 259).

<table>
<thead>
<tr>
<th>segment</th>
<th>A section of video within the Morae video file that has been designated by a researcher. It begins with an “in point” and ends with an “end point.”</th>
</tr>
</thead>
<tbody>
<tr>
<td>split-attention effect</td>
<td>Chandler and Sweller (1992) found that this learning effect is evident, when learners are required to split their attention between different source of information (e.g., text and diagrams).</td>
</tr>
<tr>
<td>worked example</td>
<td>“A worked example is a step-by-step demonstration of how to perform a task or how to solve a problem” (Clark, Nguyen, Sweller, 2006a, p. 190)</td>
</tr>
<tr>
<td>worked-example effect</td>
<td>Sweller and Cooper found learners who studied worked examples performed significantly better than learners who actively solved problems (Sweller &amp; Cooper, 1985; Cooper &amp; Sweller, 1987).</td>
</tr>
</tbody>
</table>

**Summary**

This concludes chapter 1, the introduction. This chapter outlines the purpose of this study. In particular, it explains the proposed study intends to test the “worked example effect” using animated demonstrations. This chapter posed the general research question of the dissertation. It explains that the dissertation intends to study Sweller’s worked example effect; specifically it will compare several presentation forms, animated demonstrations and discovery-practice given procedure-based learning. Finally the study’s methodology was also briefly described, in which it was mentioned that learner on-screen action will be recorded. In addition, it describes the limitations of behavior analysis and the tools of data collection. Finally it should be stated that Chapter 1 is just an introduction to the study. Chapter 2 is an extensive literature review fully describing cognitive load theory as it relates to the design of animated demonstrations. And finally, Chapter 3 fully describes the methodology of the study.
CHAPTER 2 – LITERATURE REVIEW

This dissertation argues that practicing learned procedures during early schema acquisition, increases extraneous cognitive load, resulting in decreased learner performance. This literature review lays a foundation for this argument. Specifically it reviews learning (schema and cognitive load theories), problem solving, worked examples, discovery learning, animation as an instructional strategy, and finally it concludes with an argument for this research.

Learning

Learning is like breathing. That is, human learning is a process of physiology. Learning, given this perspective, is the process of storing information as chemical changes in the brain. Physiological psychologists describe learning this way: “Learning refers to the process by which experiences change our nervous systems and hence our behavior. We refer to these changes as memories.” (Carlson, 1998, p410).

Procedural and declarative learning

Authors from a cognitive perspective make a distinction between several types of learning and memory. The two main types of learning discussed are declarative and procedural learning (Squire, 1993). Declarative learning is concerned with the learning of language-based information (e.g. facts and events), while procedural learning is skill-based learning (e.g. learning how to use a computer program) (Squire & Zola, 1996). This distinction is based upon studies involving the learning capabilities of brain injured patients, primates, and normal humans (Squire, 1986, Mishkin, 1978, Scoville & Milner, 1957).
Nearly thirty years ago, scientists studying amnesia patients published the following in the journal *Science*: “Amnesia seems to spare information that is based on rules or procedures, as contrasted with information that is data-based or declarative – ‘knowing how’ rather than ‘knowing that’” (Cohen & Squire, 1980, p.207).

Neuroscientists went on to use brain imaging techniques in the 1990s, to find that procedural learning is associated with the striatum (Squire and Zola, 1996; Poldrack & Gabrieli, 2001) while declarative learning relies on the medial temporal lobe (Grafton, Mazziotta, Presty, Friston, Frackowiak, & Phelps, 1992; Squire, 1992; Thompson & Kim, 1996; Bear, Connor, & Paradiso, 2001).

Brain anatomy and physiology may or may not seem relevant to some educators, but it is important to realize that because these two types of learning occur in different areas of the brain, they must have very different properties. As the next section will show, learning how to use software is the acquisition of procedural knowledge.

*Procedural knowledge acquisition*

Anderson’s ACT framework is perhaps the best explanation of skill acquisition (Anderson, 1993, 2005). It has changed over the past thirty years since its original conception in the 1970s (Anderson, 1976, 1983, 1993, 2005). According to Anderson’s ACT* framework (Anderson, 1983) the steps of a procedure are reinterpreted and organized into procedure-specific production rules. During this production rule stage Anderson (1983) suggests learners mentally rehearse procedures and it is even common to observe learners in this phase verbally rehearsing the steps of a procedure. Production rules involve if/then statements like the following:
IF the goal is to communicate with someone in another state

THEN dial the telephone.

A grouping of production rules is called a production. Productions are similar to what a behaviorist would have called a stimulus-response pair, but from a cognitive perspective (Anderson, 1983), because this process requires decision making and memory. This production rule stage sometimes happens after (or simultaneously) with the example generalization stage (Anderson and Finchman, 1994).

Pirolli and Anderson (1985) was perhaps the first study within the ACT framework that demonstrates the importance of examples in procedure-based learning. Perhaps because of this study, Anderson and Finchman (1994) alter the original declarative-only origin of skill learning, to include the use of examples as a means of procedural knowledge acquisition.

Anderson and Finchman (1994) suggest that examples are the only experience a novice has with the new problem category, and thus they tend to draw upon them, as they would a reference book. As they practice and experience this problem type with other examples, they produce abstract rules to help them solve similar problems. During the later stages of learning, they may generalize their learning to only remember the production rules; it is then that they are able to automate their skills, to perform a procedure as an expert.

Anderson, Finchman and Douglas (1997) complete this transition to describe a framework of four overlapping stages of skill acquisition that begins with a) an analogy stage, when learners refer to specific examples; b) later learners begin to describe abstract rules; c) then production rules d) and finally retrieval of specific examples that match the
target problem. Anderson and Finchman (1994) describe this as “proceduralization” (p. 1322) or as the “example-to-rule” method of skill acquisition (Anderson et al, 1997).

Finally according to Anderson’s most recent framework, procedural skill acquisition initially develops from example-based processing, to become rule-based processing (Anderson & Finchman, 1994). Initially specific examples may be represented in working memory, and then later, after practice, these examples are generalized into production rules (Anderson & Finchman, 1994).

Eventually as learners practice their arduous actions are automated, to be converted into the fluid movements of an expert (Schneider and Shiffrin, 1977; Shiffrin and Schneider, 1977). As the next section will show this overall process of progressing from a novice level, to that of an expert is the process of attaining expertise.

**Expertise and schema theory**

Sweller’s work relies heavily on schema theory and the differences between experts and novices. Much of the research in this area was accomplished just prior to Sweller’s 1988 cognitive load theory. For this literature review to be comprehensive, cognitive load theory must be described in context. Thus the ideas presented here lay a foundation for the rest of the dissertation.

**Information processing and human problem solving**

In the late 1950s, several computer scientists suggested that if they studied artificial intelligence, they may, one day, be able to explain human decision making processes during problem solving (Newell, Shaw, & Simon, 1958a). Simon and Newell (1971) describe their research agenda as studying the underlying processes of software to
simulate the naturally occurring processes of human neurophysiology. They reasoned that they could draw inferences from their simulations, to test and refine their theories of human problem solving. They proposed that eventually this strategy could be used to explain human behavior during problem solving (Simon & Newell, 1971).

Simon and Newell’s work began when they published a now classic paper entitled “Elements of a Theory of Human Problem Solving,” (Newell, Shaw, & Simon, 1958a). In this paper, they describe problem solving behavior as relying on information processes. They postulated that information processing systems rely on a series of primitive processes that operate in the system’s memory. These processes can be grouped together as an algorithm or “program” that describes the behavior of the system, be it a computer or a human (Newell, Shaw, & Simon, 1958b).

Throughout the 1960s, Simon, Newell and others continued their work to simulate human cognitive processes by developing a software program called “the General Problem Solver” (Ernst & Newell, 1969). They studied this software and human subjects and how they both played the game chess. Chess was studied because of the decision making processes required during an important human problem-solving process called means ends analysis (Newell & Simon, 1956; Newell et al., 1958).

Chess and human expertise

Simon continued his work on problem solving theory by studying expertise in human chess masters (Chase & Simon, 1973a, 1973b). The expertise literature of the 1980s and 1990s developed from these early studies. Simon compared chess masters to
novices in order to understand the basis of skill acquisition and expertise. In doing so he replicated a series of studies generated 30 years earlier by De Groot (1965).

Chase and Simon (1973a, 1973b) were able to determine that experts were not mentally different from novices, but that experts had recorded a vast wealth of experiences in long term memory. They determined that an expert’s memory of chess piece positions is limited to game scenarios, and not for random piece placement. Therefore the chess pieces must be in meaningful positions for a chess master to have superior recall. So in short, they found that it is this ability to perceive familiar patterns, which sets a master apart from a novice.

Simon and Gilmartin (1973) calculated that an expert must have thousands of stored chess patterns in long term memory. Chase and Simon (1973a) described these patterns or “chunks.” Miller (1956) was the first to use this term in this context, but Chase and Simon elaborated on this term to further their theories. A “chunk,” describes the amount of information being manipulated in short-term memory. Here is how they describe a “chunk” in relation to short term memory:

“Specifically, if a chess master can remember the location of 20 or more pieces on the board, but has space for only about five chunks in short-term memory, then each chunk must be composed of four or five pieces, organized in a single relational structure.” (Chase & Simon, 1973a, p.56)

Through experimentation, Chase and Simon were able to determine that experts do not have superior short term memories, just that the expert’s “chunk size” is larger. In short, experts have the ability to manipulate more information in a shorter amount of time because they recognize patterns (due to their experience in a domain).
Eventually this term “chunk” was replaced by another that had already been a part of the literature for many decades — *schema*. Schema theory is perhaps the most important component of cognitive load theory. Its origins and implications are discussed in the next few sections.

*Schema theory*

Schema theory is often credited to Sir Frederic Bartlett (1932, 1958). Even though this is the case, Rumelhart (1980) cautions us that Immanuel Kant proposed a schema theory in 1787, and that Kant’s theory more closely resembles our modern theory. Regardless of its origins, it should be stated that schema theory is supposed to account for all human knowledge, and because of this ambitious goal, it has become a somewhat complex or diverse theoretical framework.

In the 1970s, authors described schema theory in many different ways. Some described schemas as being similar to procedures (Rumelhart, 1980), or theories (Rummelhart & Norman, 1978), while others suggested they have much in common with conceptual knowledge (Bobrow & Norman, 1975; Rumelhart & Ortony, 1977).

Price and Driscoll (1997) present a more recent view and suggest that “Each schema is made up of related concepts, involving both declarative and procedural knowledge” (Price & Driscoll, 1997, p.476). In short schemas are data structures within long-term memory that are related to concepts or patterns of behavior (Rumelhart, 1980).

*Schemas as problem categories*

Simon continued his work to simulate human thought processes, and in the mid 1970s studied learners as they solved algebra problems (Hinsley, Hayes, & Simon, 1976).
Simon and his associates reasoned that if learners used schemas to understand and interpret verbal information, they may also use them to categorize problems. Through several experiments, they found that indeed both computers and humans tend to categorize problems during problem solving, and more importantly, use their memory of problem categories to solve these problems.

Chi, Feltovich, and Glaser (1981) built on Simon’s work to relate schema theory to expertise and problem solving. While studying physicists, they found that these experts gather information to categorize problems, and once a problem is categorized, the expert uses a set of schema specific production rules to solve that problem.

Sweller (1988) uses a similar definition to describe problem schemas and describes them in terms of problem representations or problem states. “These cognitive structures will be called schemas where a schema is defined as a structure which allows problem solvers to recognize a problem state as belonging to a particular category of problem states that normally require particular moves.” (Sweller, 1988, p. 259).

Sweller’s work in many ways, follows the same line of reasoning as that of Simon. Sweller (1988) also uses a computational model to suggest that learners learn how to do algebra problems by learning problem configurations, just as chess experts learn board configurations. This then is the context within which Sweller developed cognitive load theory. Sweller (1988) suggest cognitive load theory was developed as a means of explaining expertise and the performance of novices during schema acquisition.

*Cognitive load theory*

Cognitive load theory proposes that because working memory is limited, learners may be bombarded by information during the initial stages of learning, and if not
properly managed, this cognitive overload can prevent learning and eventually deteriorate performance (Sweller, 1988). Thus this working memory load has important consequences for learning and the design of instruction (Sweller, 1993, 1999).

Sweller and his colleagues describe three different types of cognitive load -- intrinsic, extraneous and germane cognitive load (Sweller, Van Merriënboer, & Paas, 1998; Paas, Tuovinen, Tabbers, & Van Gerven, 2003).

**Figure 1. Cognitive load over time (adapted from Paas et al, 2003)**

*Intrinsic cognitive load* is the inherent level of difficulty (or complexity) associated with an instructional activity. *Extraneous cognitive load* is that load not inherent within the activity (Chandler & Sweller, 1991; Chandler & Sweller, 1992), but controlled by Instructional designers as they structure and present instructional materials.
(Pollock, Chandler, & Sweller, 2002). Finally *Germane (or Relevant) cognitive load* is that remaining free capacity, which may directed toward schema acquisition (Sweller et al., 1998).

*Element interactivity*

Intrinsic cognitive load is due to the inherent complexity of instructional materials. It is what separates a calculus problem from a second grade math problem. Sweller and Chandler (1994) suggest the level of complexity of each problem is due to the number of interacting elements within the problems; they describe this phenomenon as *element interactivity*. According to Sweller and Chandler, instructional content is composed of component parts or “elements;” elements may be said to “interact” if there is a relationship between them, thus raising the complexity of the instruction. Van Merriënboer and Sweller (2005), suggest the total number of elements is not as important as the number of interactions between the elements.

Van Merriënboer and Sweller (2005) clearly describe this problem when they mention “Working memory must inevitably be limited in capacity when dealing with novel, unorganized information because as the number of elements that needs to be organized increases linearly, the number of possible combinations increases exponentially” (Van Merriënboer & Sweller, 2005, p.149). In addition, Sweller and Chandler argue that the intrinsic cognitive load can not be changed by the Instructional Designer, specifically they state “Elements may interact either because of the intrinsic structure of the information or because of the manner in which it is presented or both. The intrinsic structure of information is unalterable...” (Sweller & Chandler, 1994, p.26).
While this may be the case, instructional designers still have the ability to manage cognitive load.

Managing cognitive load

Sweller et al (1998) describe not one but three types of cognitive load. Even though Sweller and Chandler suggest that the intrinsic load of instruction is “unalterable,” they have empirical evidence that Instructional designers can lower the extraneous load of instruction, by altering the presentation of those materials (Sweller & Chandler, 1994).

Sweller and his associates argue that even when the cognitive load of instruction is very high, instructional designers can artificially reduce the intrinsic load of instruction, by dividing a lesson into smaller pieces, thus reducing the intrinsic load of the overall lesson. Sweller describes these smaller pieces as “subschemas” (Clark, Nguyen, & Sweller, 2006b). This method of presentation was first developed by Pollock, Chandler, and Sweller (2002).

This method of dividing a lesson into subschemas promotes learning at the expense of understanding, but as Sweller explains they are unable to understand the full schema anyway (Clark, Nguyen, & Sweller, 2006b). They found that if learners process the individual elements of instruction serially, rather than simultaneously, that they were better able to process that instruction. It should be noted these researchers were not the first to suggest breaking instructional materials down into its component parts. Gagné recognized this in the 1960s (Gagné & Paradise, 1961; Gagné, 1968). But it is important to realize that Sweller and his associates not only recommend this method of instruction,
but were also able to explain why Gagné’s learning hierarchies are an effective means of presenting instruction.

Problem solving and cognitive load

While problem solving skills are highly valued, many problems are complex cognitive tasks that may be difficult for a novice to complete, even when they have the prerequisite skills (Sweller, 1988; van Merriënboer, 1997). Complex cognitive tasks require learners to mentally reorganize what they know, to restructure the problem, in order to accomplish the overall task (van Merriënboer, 1997). This mental reorganization may impose a high cognitive load on the learner (Sweller, 1988), but this load may be manipulated by an instructional designer during the design of instruction, to allow a learner to learn (Sweller, 1993).

Means-ends analysis and cognitive load

Problem solving has been studied for many decades, Newel and Simon studied problem solving as early as the mid 1950s (Newell, Shaw, & Simon, 1958a). As described in previous sections, much of cognitive science and schema theory developed out of their work. Both Newel and Simon, and Sweller suggest that novices solve problems by using an iterative process called “means ends analysis.’’

During means ends analysis the learner applies problem solving operators to achieve a sub goal of the problem, once this sub goal has been reached, the learner then reassesses the problem, to continue to apply other problem operators until the problem goal is reached. Chi et al. (1981) found that experts work somewhat differently, in that they categorize a problem, based upon the deep structure of the problem, to work forward
toward a problem solution. But novices unaware of the problem schema must use a means-ends analysis approach to work backwards from the problem goal (Larkin et al., 1980).

Ward and Sweller (1990) describe the use of this means ends analysis strategy this way: “A heavy cognitive load is imposed because of the need to simultaneously consider and make decisions about the current problem state, the goal state, differences between states, and problem solving operators that can be used to reduce such differences” (Ward & Sweller, 1990, p.3).

Unfortunately some learners are required to solve problems when they are not truly aware of the underlying problem schema. If a learner is required to solve problems before they understand the problem schema, they may become distracted with irrelevant aspects of a problem, spending their time searching for a problem solution, but still may not be engaged in learning (schema acquisition) (Sweller et al, 1998).

So in the case of photo editing, a novice edits a document using the software interface (using problem solving operators), until they produce the desired product (problem goal). But when novices are learning how to use graphic design tools for the first time, they usually have difficulty and make mistakes. The actions of an expert graphic artist are much more rapid and precise because they work forward with a plan in mind. So even though a novice may have a problem goal in mind, and may know how to use the tools, they may not be fully aware of how to produce the problem goal.
Worked examples

Working memory is limited (Baddeley, 1986; Miller, 1956; Peterson & Peterson, 1959), therefore a novice’s attentional resources are limited (Sweller, 1998) thus novices may become distracted by irrelevant aspects of a problem, and make errors during problem solving. Later, because of these errors, they may have difficulty recalling or piecing together the underlying steps of the overall procedure. So rather than trying to solve problems initially, Sweller and Cooper (1985) suggest learners should use that time more effectively, to first, learn the underlying problem schema, perhaps by studying examples, and then once learners understand the problem schema, then practice to automate their skills. Thus Sweller and his associates began to research worked examples.

In the mid 1980s, Sweller and Cooper compared learners who studied worked examples, to those learning by traditional problem solving. In a series of five experiments, Sweller and Cooper (1985) measured the performance of high school and undergraduate learners as they learned algebra problems. They found that those students who studied worked examples took less time to process the instructional materials, and subsequently took less time to solve problems. In addition they also had a decrease in mathematical errors. This phenomenon has subsequently been termed the worked-example effect (Sweller et al, 1998). Cooper and Sweller (1987) replicated their earlier findings, to find that learners using worked examples also spent less time solving transfer problems and that they made significantly fewer errors on transfer problems.

In 1988, Sweller developed cognitive load theory to explain novice behavior during early schema acquisition. Owen and Sweller (1985) had shown that learners, that
solved problems by means ends analysis, would have a higher working memory load as compared with those who were prevented from using a means ends strategy. Sweller and Cooper (1985) used worked examples to limit means-ends analysis during schema acquisition. Cooper and Sweller, (1987) suggest this instructional strategy was designed to limit problem solving search, in order to alleviate the cognitive load imposed on a novice. Cooper and Sweller found that by removing problem solving search, that learners were more efficient, and made fewer problem solving errors.

So not only did Sweller provide evidence for his theory (Sweller & Cooper, 1985; Cooper & Sweller, 1987), but also a mechanism for why problem solving is less efficient (Sweller, 1988). This mechanism, problem solving search, is removed from worked examples.

*When are worked examples warranted?*

Recall that Sweller’s cognitive load theory was designed to explain the behavior of novices as they solve problems (Sweller, 1988). While worked-examples are useful for novices, they must eventually practice a procedure to attain expertise. Worked examples then are only warranted for novices (Kalyuga, Chandler, Tuovinen & Sweller, 2001), thus this instructional technique only serves as an introduction for learners during early schema acquisition.

Interestingly enough, Kalyuga et al (2001) found worked examples actually hindered more advanced learners, because once a skill has been acquired, the worked examples become redundant and overload the working memory of experts (Kalyuga et al,
2001). This has since been termed the expertise reversal effect (Kalyuga, Ayres, Chandler, & Sweller, 2003).

Thus, as learners become more competent, the worked-example and other cognitive load effects disappear (Kalyuga, Chandler, & Sweller, 1998). Recall that Anderson’s ACT-R framework suggests learners learn production rules later in the learning process. Thus ACT-R predicts a gradual reduction of the effects of cognitive load. Once an expert automates these skills, worked examples become unnecessary.

Chandler and Sweller (1991) suggest that extraneous cognitive load is due to the format of the instruction. In other words the presentation strategies employed during the design of instructional materials can cause learners to perform poorly. Sweller and his associates found other cognitive load effects due to the presentation techniques employed. The next couple of sections introduce several of these learning effects. What’s important to realize from this discussion is that all of these effects can be applied to worked examples, which like any instruction can overload the learner if not properly designed (Ward & Sweller, 1990).

The design of effective worked examples

Several researchers have devised presentation strategies for producing more effective worked examples that help learners abstract a problem schema. Specifically this section describes cognitive load theory as it relates to problem structure and later the discussion will turn to problem format and the use of multimedia.
Problem completion

Van Merriënboer (1990) followed up on Sweller’s early studies to describe another useful strategy for reducing cognitive load. Usually cognitive load is reduced by presenting worked examples as instruction, along with problems for the learner to practice and solve (Sweller, 2003). But van Merriënboer (1990) suggested that it may also be useful to provide learners with partially worked examples and for those learners complete these problems. In doing so, it is expected that these learners will study the worked-out portion of the problem, to abstract that component of the problem schema, providing them with guidance while lessening their cognitive load, but still allowing them some practice.

Van Merriënboer and de Croock (1992) compared this problem completion strategy with learners who were actively engaged in more traditional problem solving. Their study involved learners enrolled in an introductory software programming course. They gave one group of learners a library of partially completed computer programs, and then measured their performance versus another group who had to write their programs on their own. Van Merriënboer and de Croock found that “problem completers” were more successful on both program construction tests and multiple choice tests. This finding was later replicated by Paas (1992), but he also found that problem completers are better able to transfer their learning to new situations. This was subsequently named the problem-completion effect (Paas, 1992).

Thus the problem completion strategy provides learners with a problem statement and only a partial solution, while the worked example strategy provides learners with a problem and the entire problem solution. Each of these strategies helps learners to limit
their cognitive load and to refocus their attention on schema related material, thus promoting germane cognitive load.

Sweller, van Merriënboer, and Paas (1998) published an important article, describing the benefits of germane cognitive load. In doing so this article has begun to refocus this literature, from only concentrating on decreasing extraneous cognitive load, to now considering the idea of decreasing extraneous load to redirect attention toward germane load.

*The variability of worked examples*

Paas and van Merriënboer later turned their attention toward worked examples and in 1994 and were among the first to describe yet another cognitive load learning effect, *the variability effect*. This effect suggests that the variability of worked examples encountered during instruction, is important. This is because learners are better able to abstract a problem schema, if they are provided with multiple examples of that problem type. This reasoning is in line with Anderson’s “example-to-rule” method of skill acquisition (Anderson et al, 1997).

Paas and Van Merriënboer found that if learners are exposed to a broad variety of problem types during instruction, that they are later, better able to recall the underlying problem schema. In their words:

“*Results showed that students who studied worked examples gained most from high-variability examples, invested less time and mental effort in practice, and attained better and less effort-demanding transfer performance than students*
who first attempted to solve conventional problems and then studied work examples.” (Paas & van Merriënboer, 1994, p.122)

While it is important that learners be provided with a variety of examples, their last point is also important. For this study also shows that it is better to study worked examples first, and then later practice those problems for better retention.

Quilici and Mayer (1996) used a series of experiments for a partial replication of Paas and Van Merriënboer’s 1994 findings. Quilici and Mayer studied learners who had studied statistics examples (e.g., t-test, chi-square, and correlation problems) and found that those learners who were asked to categorize worked examples were later more likely to sort problems on the basis of problem structure than on the basis of surface characteristics.

Recall that Chi, Feltovich, and Glaser (1981) suggested that experts differ from novices, in their ability to distinguish between problem categories. This is the basis of our modern schema theory. Quilici and Mayer’s results demonstrate that learners who have studied worked examples have been trained to act more like experts, in that they categorize problems like experts.

Split-attention

Tarmizi and Sweller (1988) noticed that the worked example effect did not work for all worked examples. In several studies, they studied learners who used traditional diagrams versus those who used integrated diagrams (like those in figure 2).

They found learners who used integrated diagrams, were better able to process that information (Ward & Sweller, 1990; Chandler & Sweller, 1991; Chandler & Sweller,
1992). Or more specifically that if text (a visual form of instruction) is simultaneously presented to the learner with a diagram (also visual instruction) there is a potential for overload. This phenomenon was later described as the *split-attention effect* (Chandler & Sweller, 1992).

**Figure 2. Split-Attention Diagrams**

Sweller suggests that while structuring materials, instructional designers must be careful how they direct learners attention within that instruction (Ward & Sweller, 1990). Thus even worked examples can be ineffective if they raise a learner’s cognitive load to overload levels (Ward & Sweller, 1990).

This split-attention effect is not limited to geometry, Chandler and Sweller (1991) found that this effect extends to a variety of other disciplines and is simply a limitation of human information processing. Since information can be encoded both as text (visually) and narration (auditory), split-attention is a potential problem that exists within animated demonstrations as they would in any worked examples (Tarmizi & Sweller, 1988). So,
Instructional designers need to remove text from animated demonstrations, to limit any possible split attention. In addition narrated auditory versions of that text should be used if at all possible.

The split-attention effect is an important example of how presentation techniques can alter learning. But this is learning effect is limited to single modality instruction. In the 1990 several researcher began to study multimodal instruction. This was inevitable given the ubiquity of the personal computer and CDROM in the early 1990s. But many of these studies considered cognitive load as they compared visual only conditions, or combinations of text and animation. The next section describes this literature as it relates to cognitive load theory.

**Cognitive load and multimodal instruction**

Cognitive load theory relies on the human cognitive architecture developed in the 1960s by Atkinson and Shiffrin (Kirshner, Sweller, & Clark, 2006). This architecture divides human memory into three subsystems, a sensory memory, a short term store (or working memory), and a long term store (Atkinson & Shiffrin, 1968). Baddeley further subdivided working memory into two separate visual and auditory subsystems, and made it clear that these subsystems operate independently of one another (Baddeley & Hitch, 1974; Baddeley, 1986). This basic plan was also proposed a few years earlier by Paivio as the dual coding hypothesis, and then later as dual coding theory (Paivio, 1971, Paivio, 1978).

Clark and Paivio (1991) later suggested that the dual nature of working memory is important for those designing instruction. Following this several researchers found
empirical evidence that suggested learner performance could be improved when using multimodal instruction. In a series of experiments these researchers found that learners working with multimedia consistently out-performed those learning with text-based materials (Jeung, Chandler, & Sweller, 1997; Mayer & Anderson, 1991; Mayer & Moreno, 1998; Mousavi, Low, & Sweller, 1995). This has since been described as the modality effect (Moreno & Mayer, 1999; Mousavi, Low, & Sweller, 1995; Penney, 1989) or the Modality principle (Mayer, 2001).

Mousavi, Low and Sweller (1995) were perhaps the first to provide an explanation for this modality effect. This is because they considered dual-modality presentations from a cognitive load perspective. They found that under high load conditions if an instructional designer moves the instructional message from a visual mode (text) to an auditory mode (narration) that this increases learner performance.

Mousavi, Low and Sweller (1995) reasoned that when a lesson is structured so that it uses both modalities, learners are able to use both working memory subsystems simultaneously to reduce their overall cognitive load by distributing that load to these independent subsystems. Specifically they suggested that performance is improved because multimodal instruction increases the learners’ “effective working memory capacity” (Mousavi et al, 1995, p319).

The mode of instruction

The mode of instruction depends on the material. While multimodal instruction is useful, Clark, Nyguen, and Sweller (2006b) suggests some material is best presented in verbal form and others in visual form. A combination is useful to reduce cognitive load,
but Sweller suggests Instructional designers should usually always represent a square in
visual form (Clark et al., 2006b). People can quickly identify this graphic information if it
is presented visually, while they will usually always take longer to identify it, if it is
presented verbally (e.g. one side is vertical to the others at a 90 degree angle, while the
next is, etc.). Alternatively the concept of truth is very difficult to represent visually.
Clark (1999) explains the difference between these two concept when she describes them
as concrete and abstract.

In the end, most would agree with the idea that some material is best presented
either visually or verbally depending on the materials. The next section follows this line
of reasoning to discuss animation as an instructional medium.

*Animation as an instructional strategy*

Animation is often used in a gratuitous manner, with the decision to use it more
for marketing purposes, rather than for reasons of instructional effectiveness (Rieber,
1990, 2000). A number of articles have been published which describe the effectiveness
of animation as a presentation technique. This section introduces this discussion and then
turns to the animated demonstration literature.

*Instructional uses of animation*

Rieber (1990) reviews the literature concerning animation. At that point few if
any empirical investigations had been performed given animated instruction. Rieber
(2000) updates this review and made several suggestions concerning animation as an
instructional medium.
Rieber suggest studies before 1970 concluded that graphics did not aid and could even distract learners, but following this date evidence began to mount in favor of the use of visuals to support learners. Rieber (1990) reviews twelve studies concerning learner performance that were mixed in their results. However, from these studies he was able to draw some conclusions and provide several useful guidelines concerning the use of animation in instructional materials.

Specifically Rieber (2000) suggests two important guidelines concerning the use of animation. First like all graphics, animated instruction must “pass the test for a need for ‘external visualization’” (p. 162). This is a lesson learned from the static graphics research. Rieber then suggests that many of these lessons can be applied to animation. But animation and static graphics do have their differences. Specifically, Rieber (2000) suggests the use of animation, when the learning of content requires changes in an object motion or trajectory, or both. He suggests that learning effects may be greater if both motion and trajectory changes are a part of the instructional materials be it for procedure, concept, or principle-based learning.

Finally Rieber (1990) makes a very important point, that animation should be included only when the attributes of the animations match the task. This last point, instruction that matches the task, is one that is violated by the best of us.

*Animated Instruction that matches the task*

Rieber and Parmley (1992, 1995) developed animated instructional materials, to teach learners the basic principles of Newtonian physics (specifically the laws of motion). In this case they use interactive animation (or simulation) to teach learners how to control
a simulated space shuttle, given structured and unstructured lessons. Their structured lessons allowed learners increasing levels of control, that is new “subskills” were taught during successive levels of instruction. They compared this “structured” tutorial to unstructured activity (discovery-based learning) with full control from the beginning. The result was that there were no significant differences in their conditions.

Even though Rieber and Parmley’s tutorial was in part procedure-based, it also had a conceptual or principle-based component. Sometimes in complex domains like physics, it is difficult to separate the two. In this case however, Rieber and Parmley (1992) measured outcome measures that were primarily principle-based, while their instruction was primarily procedure-based. It is understandable that Rieber and Parmley wanted to teach learners physics principles given simulation, but perhaps they should have chosen another medium or another type of animation. Like many researchers they compared groups of learners, given their performance on a multiple choice test (in a pretest/posttest design). This use of animation was innovative, but procedure-based learning requires observational data collection.

Animated demonstration is a form of animated instruction that has a main purpose, to teach learners how to perform procedures. This form of instruction may be used to teach learners other types of learning (e.g. concepts) but this may be a misuse of the medium. This study will endeavor to use this form of instruction properly and to use observational data to measure learner performance.

This discussion implies an important distinction for this dissertation project. Animation may primarily be used to teach learners procedures (animated
demonstrations). But it can also be used to teach learners conceptual material (animated explanation).

*Animated explanation and animated demonstration*

It is thought that animated demonstrations act as animated worked-examples (Lewis, 2005). Sweller and his colleagues even define worked examples as a form of demonstration as they describe them this way: “A worked example is a step-by-step demonstration of how to perform a task or how to solve a problem” (Clark, Nguyen, Sweller, 2006, p. 190). While this is the case little to no cognitive load research has been conducted using animated demonstrations.

Some would quickly dismiss this statement to begin describing the work of Richard Mayer, for Mayer and his colleagues have been quite prolific over the last decade. Mayer and his colleagues have indeed used animation extensively in their instructional conditions and are well known for their contributions to the modality effect literature. But the instructional materials used in this literature, is more aptly described as animated explanation rather than animated demonstration.

Mayer describes his work well when he explains that his experiments ask questions about scientific explanation: “By ‘explanation’ we mean a description of a causal system containing parts that interact in a coherent way, such as a description of how a pump works or how the human respiratory system works.” (Mayer & Sims, 1994, p.389). Clark and Mayer (2003) even describe these as “two different e-learning goals” that teach learners to “inform and perform “(p.17). Thus this dissertation argues learning “how to do” something (perform) is best taught via animated demonstration. As opposed
to teaching a learner about something (inform) which is best taught via animated explanation.

Interestingly enough, Mayer’s studies with animated explanations only found significant differences between instructional conditions given transfer. Other researchers (e.g. Cooper & Sweller, 1987) have found significant differences in other important outcome variables such as completion times and the number of errors. Given Anderson’s well respected ACT-R theory, it is likely to expect that these two forms of learning are dramatically different.

This study argues that animated explanations and demonstrations are clearly different because of the type of learning that occurs (declarative and non-declarative learning; Squire). Animated demonstration represents a more cognitively demanding form of learning from animation. This is because by its very nature, demonstration assumes the eventual action of the learner. So rather than just encoding information that describes a system, the learner is encoding rules based upon a sequence of actions that they will have to perform. Performance is when the true cognitive load of the situation is highest, for it is then that the learner will have to recall the actions taught during the animation (if, then rules), but it then that they will also apply those rules, in sequence, to produce the goal of the current situation. It is also then that learners need well constructed instructional materials that reduce the extraneous cognitive load imposed by problem solving. This then leads us to the animated demonstration literature which is of primary importance to this literature review.
Animated demonstration

The animated demonstration literature extends back to the early 1990s, however this literature was written before much of the modality, or cognitive load literature existed. Thus this section will concentrate on placing this form of instruction in context with more recent instructional design literature. Specifically it considers animated demonstrations given the modality and split-attention effects.

The animated demonstration literature is at times complex and contradictory (please note Appendix E). This is because an animated demonstration may be produced with or without audio, and they may or may not include text annotations. Thus there are studies comparing animated demonstrations using each of these types of media and combinations of these media.

Animated demonstrations in the early 1990s

During the early 1990s, the World Wide Web was in its infancy and much of the animation of the day was presented via personal computers, mainly via CDROM. In addition many investigators used a Macintosh hypertext environment known as HyperCard.

Waterson and O’Malley (1993) is a good example of an early 1990s animated demonstration study. At that time researchers were very interested in comparing “media effects,” recall this was the height of the famous Clark-Kozma debates (Clark, 1983; Clark, 1994; Kozma, 1991; Kozma, 1994). In the past decade researchers have begun to look at learning from a cognitive load perspective and can now reinterpret these earlier
Waterson and O’Malley’s findings.

Waterson and O’Malley (1993) evaluated the effectiveness of several forms of animated demonstrations given a set of six discrete tasks. Their instructional conditions included animated with text, animation only (no text or narration), and a combination group (animated demonstration with text and narration). Participants were taught a Macintosh graphing application called Cricket Graph. This instruction was presented via HyperCard.

Waterson and O’Malley measured performance time given three instructional conditions, with three types of tasks, identical, similar, or different tasks from those that learners had initially learned. The data revealed a significant main effect with respect to group. The combination text-narration group outperformed the text only, and no narration groups (Note Figure 3, a graph of the performance times of their participants).

Waterson and O’Malley’s data also revealed a significant interaction between type of instruction and task group as the combination group completed tasks sooner than the other groups. They also found a trend in the data that suggested learners using the text only instruction were slower than either of the other groups (animation only and animation with text and narration).
The instructional conditions in this study show some interesting cognitive load effects. The fact that the slowest performance times were from the text-based animated demonstration group is evidence that the split-attention effect can negatively affect learner performance given animated demonstration.

It’s also interesting that the combination group had decreased performance times. This is evidence for the modality effect given animated demonstrations. The fact that the combination group had a text redundancy for the narrated message may be little reason for concern, since this group out-performed the other groups. However it is possible that learner performance may still be increased by removing this redundancy. Interestingly enough these learners may have ignored this redundancy to have benefited from the modality effect.
Even though Waterson and O’Malley conducted a series of repeated measures ANOVAs, demonstrating that animated demonstrations show some promise, the study unfortunately only had 30 participants (10 per condition). This perhaps is sufficient for a pilot study to test the instruments, but the results of this study are somewhat suspect due to a lack of power.

*Palmiter’s animation deficit*

Palmiter’s 1991 dissertation project is perhaps the most cited animated demonstration study. In a series of experiments she compares the learner performance of those who study animated demonstrations with those that use text-based instructions. In a repeated measures study, she studies learners as they completed a set of discrete HyperCard tasks during a training session, an initial test, and a delayed test. This study measures four dependent variables – performance time, accuracy, retention and transfer (Palmiter, 1991).

Palmiter (1991) found several significant session x media interactions. These results are quite interesting for she found that during the training session learners who study animated demonstrations sped skill acquisition and performed tasks in less time and more accurately than their peers using text-based instruction (Palmiter, Elkerton, & Baggett, 1991). These results are similar to those by Sweller and Cooper (1985) who demonstrate that studying worked examples requires “considerably less time to process than conventional problems, but that subsequent problems similar to the initial ones also were solved more rapidly” (p.59).
However, Palmiter (1991) noted that one week later, learners using animated demonstrations took longer to perform tasks as compared with learners using text-based instruction. In short their skill acquisition may have been quicker during the initial training session but their retention was lacking one week later. This phenomenon, later described as an “animation deficit” (Lipps, Trafton, & Gray, 1998) could not be replicated by either Waterson and O’Malley (1993) or Lipps et al (1998).

Palmiter et al. (1991) also measures transfer. She measured performance time while performing tasks similar to those trained (e.g. Copy Button vs. Copy Field). Demonstration groups took significantly less time than the text only groups during the “immediate test session.” This again is very similar to results by Cooper and Sweller (1987) who write: “The results indicated that subjects whose training included a heavy emphasis on worked examples or an extended acquisition period were better able to solve both similar and transfer problems than were those subjects trained with conventional problems” (Cooper and Sweller, 1987, p 347). But again, Palmiter (1991) notes there was a significant increase in performance time for the demonstration groups between the training and delay sessions.

Since Palmiter’s early study, many researchers have evaluated this presentation form (Cornett, 1993; Harrison, 1995; Hegarty, Kriz, & Cate, 2003; Reimann & Neubert, 2000). But these researchers do not describe animated demonstrations as animated worked examples nor do they describe a phenomenon similar to Palmiter’s animation deficit. Quite the opposite in fact, they found learners using narrated presentations were faster and more accurate than those using text-based instruction of instruction (Cornett, 1993; Harrison, 1995).
Furthermore, none of these animation studies, compared learner performance with discovery learning. Only Rieber and Parmley (1992, 1995) compare the learner performance of those using an animation condition to using a discovery learning condition, but this study uses simulations (dynamic animation) rather than animated demonstrations.

Finally, it makes sense to separate complex cognitive tasks into discrete components. This is what Gagné (1968) described as learning hierarchies, and suggests this is how Instructional designers should design instruction. But research design and instructional design are separate activities. When assessing learning, researchers need to be cautious to maintain external validity. Real problems are not a series of discrete tasks, as in the Palmiter or the Waterson and O’Malley studies. Learners usually have to mentally plan and organize a series of discrete skills to accomplish a larger complex cognitive task. This planning and management of known skills can cause a cognitive overload, causing learners to be unable to accomplish a task or to perform them at a much slower pace.

**Conclusions about animated instruction**

Reiber (1990) suggests the use of animation in instruction is relatively new, and has only been made available given computer-based instruction. Reiber (1990, 2000) suggested it has often been used in a gratuitous manner and is only useful under a certain range of conditions. Like all graphics, animated instruction must “pass the test for a need for ‘external visualization’” (Reiber, 2000, p. 162). In addition, Rieber (1990) also suggests animation should be included only when the attributes of the animations match the task.
Early animated demonstration researchers were very interested in “media effects”. For instance, Waterson and O’Malley studied how learners learned given text, animation and narration. Since this study was first published researchers have begun to follow Richard Mayer’s advice to base learner performance on characteristics of the learner rather than characteristics of the media (Mayer, 1997).

Perhaps the most notable finding from the animated demonstration literature was the potential for an animation deficit (Palmiter, 1991; Lipps et al, 1998). As Palmiter describes this deficit, retention given animated demonstrations is a short term phenomenon, under which learners using animated demonstration were found to speed skill acquisition and performance time, in the acquisition phase, but one week later these same learners had difficulty retaining those skills, and subsequently their performance one week later was found to be lacking (Palmiter, 1991) or not significantly different (Lipps et al, 1998).

Thus two main questions are apparent given the animated demonstration literature. First Palmiter’s animation deficit is the most salient. Secondly while Lewis (2005) argues that animated demonstrations are animated worked examples, this has yet to be demonstrated empirically. Waterson and O’Malley’s results showed that the modality effect could be apparent given this form of instruction, but it can not be said that the worked example effect has been demonstrated given animated demonstrations. To do this, one would have to demonstrate that learners using animated demonstrations perform significantly better than those learning through solving the equivalent problems.
Those suggesting learners solve problems as a means of learning are using an instructional strategy first developed in the 1960s. Thus the next section considers discovery learning, and the worked example effect in relation to discovery learning.

**Discovery learning**

Jerome Bruner initiated the discovery learning movement with his 1961 article in the *Harvard Educational Review*. Bruner’s discovery learning approach is summarized well by the following passage:

“It is, if you will, a necessary condition for learning the variety of techniques of problem solving, of transforming information for better use, indeed for learning how to go about the very task of learning. Practice in discovering for oneself teaches one to acquire information in a way that makes that information more readily viable in problem solving” (Bruner, 1961, p.26).

While Bruner eloquently expresses the importance of discovery learning, he does not define discovery learning.

*What is discovery learning?*

Klahar and Nigam (2004) report that for almost 100 years that the discovery learning literature has had a consistent problem defining *discovery learning*. Authors sometimes provide definitions like the following: “…learning by discovery is defined usually as teaching an association, a concept, or rule which involves ‘discovery’ of the association, concept, or rule” (Glaser, 1966, p.14). This circular definition assumes the reader has an understanding of all these terms. However, it does gives some indication that “learning by discovery” is an instructional strategy. While this may be the case,
many others suggest discovery learning is something accomplished, by autonomous
learners (e.g. Gagné, 1966).

Glaser (1966) contrasts discovery learning with traditional instruction to suggest
that one of the most important characteristics of discovery learning is that it makes use of
induction during the process of learning. However, Wittrock (1966) explains induction is
not a prerequisite for discovery learning, that it is equally possible for a learner to begin
with a higher order generalization, to discover specific conclusions, as it is to discover
generalizations or rules from specific examples, or in his words “Induction has no
exclusive identity with discovery learning” (Wittrock, 1966, p43).

*Learning to discover*

During the 1960s there were many proponents of discovery learning, but Ausubel
(1962) was one of the earliest critics of Bruner’s instructional techniques. Ausubel
devotes a sizable portion of his early cognitive texts to this instructional strategy
(Ausubel, 1962, 1968). He suggest that there is a time and place for discovery learning in
the classroom, but that it is a highly inefficient means of conveying large amounts of
information. He suggests learners must learn vast amounts of information in their
lifetimes, more than they could ever discover on their own. Thus he finally concludes that
discovery learning “is as unfeasible as it is unnecessary” (Ausubel, 1962, p.151).

Later however Ausubel admits that teaching one to discover knowledge is an
admirable goal, but unlike Bruner (1961) who suggests it is “a necessary condition for
learning” (p.26), Ausubel writes:
“The development of problem solving ability is of course, a legitimate and significant educational objective in its own right. Hence it is highly defensible to utilize a certain proportion of classroom time in developing appreciation of facility in the use of scientific methods of inquiry and of other empirical inductive, and deductive problem solving procedures. But this is a far cry from advocating that the enhancement of problem solving ability is the major function of the school” (Ausubel, 1962, p154).

So when is it appropriate to use discovery problem solving? Many authors believe that it is unconscionable to only use direct methods of instruction that is at some point we must teach learners how to teach themselves. Cognitive load researchers are not against teaching learners to teach themselves, they are only concerned with how one introduces learners to problems and problem solving skills.

Why is discovery learning ineffective?

Given the expertise reversal effect (Kalyuga, Ayres, Chandler, & Sweller, 2003) suggest direct instruction is only useful during the earliest stages of learning, during early schema acquisition. Thus these and other researchers suggest fading worked examples, following schema acquisition, to allow more advanced learners time to practice and automate their problem solving skills (Renkl, Atkinson, Maier, & Staley, 2002). Thus it is not practice or problem solving that cognitive load researchers are against, it is just the timing of all that practice which is under scrutiny.

Klahr and Nigam (2004) suggest learners learning under discovery learning conditions are more likely to “encounter misleading feedback, to make encoding errors
and causal misattributions” (p.661). Gagné (1966) was one of the first to consider these ideas. He describes discovery learning as involving “(1) a process of search, and (2) a process of selection, each of which takes place within the learner’s nervous system” (Gagné, 1966, p. 136).

Recall that Sweller (1988) also considers problem solving search, and suggests that if a learner is required to solve problems while learning, they may spend many hours searching for a problem solution and still not be engaged in schema acquisition (Sweller, 1988, Kirschner, Sweller, & Clark, 2006). In other words, even though they are actively searching for a problem solution, they may not be learning (Sweller, 1988).

Discovery learning and constructivism

Discovery learning has morphed and changed over the decades since Bruner’s 1961 article. Mayer (2004) suggests it is still with us in the writings and practices of constructivism. During the 1990s many American educators adopted a constructivist epistemology toward teaching and learning. This epistemology strongly suggests the active construction of knowledge (Dewey, 1916; Wittrock, 1974, Duffy & Cunningham, 1996). Constructivism is primarily a philosophical position, but one that has implications for instruction. It suggests we perceive information from the environment, and that our mental models of that environment help us to construct our own unique version of reality (Jonassen, 1991).

This relativist epistemology extends into a philosophy of instructional design. Jonassen (1991) suggests instructional designers should focus less on “prescribing a single best sequence of learning (p.12)” to allow learners to negotiate their own learning.
Thus many constructivists design ill-structured learning environments because they feel the learner will construct his or her own interpretation of that environment, and must be allowed to do so.

Jonassen (2002) has updated constructivism to describe “learning as activity.” He and his colleagues are now attempting to integrate Activity Theory and Ecological Psychology into their constructivist philosophy of learning. This “learning as activity” mantra, in instructional design terms becomes “learning by doing.”

*Learning by doing*

“Learning by doing” has been a popular approach for the last half century, but more recently some Educational psychologists and Instructional design researchers have begun to question this type of instruction. For instance, Mayer (2004) describes the learning outcomes given constructivist-based instruction, as the result of “the constructivist teaching fallacy” (Mayer, 2004, p.15). Specifically Mayer suggests that most constructivists prescribe active learning techniques, which require learners to be behaviorally active during early schema acquisition.

Rather than being behaviorally active, Mayer (2004) suggests the promotion of cognitive activity. Mayer puts it best when he says “Activity may help promote meaningful learning, but instead of behavioral activity per se (e.g., hands-on activity, discussion, and free exploration), the kind of activity that really promotes meaningful learning is cognitive activity (e.g., selecting, organizing, and integrating knowledge)” (Mayer, 2004, p.17). While Mayer’s 2004 article stops short of condemning pure discovery or constructivism, it does conclude “The research in this brief review shows
that the formula constructivism = hands-on activity is a formula for educational disaster (Mayer, 2004, p. 17).”

While Mayer (2004) and Kirschner, Sweller, and Clark (2006) are very critical of constructivism and describe its instructional design prescriptions as “unguided” or “minimally guided” instruction, this epistemology is very popular in American education and has a strong following.

It seems the real problem given this debate is more one of scale and communication. Researchers need to be specific about which learners and what one means by “learning.” Kirschner, Sweller, and Clark (2006) agree with the constructivists when they say that “knowledge is constructed by learners” (p.78). That is they agree with them generally, that lower order skills can be used constructively to solve problems. But they go on to say that “The constructivist description of learning is accurate, but the instructional consequences suggested by constructivists do not necessarily follow” (p. 78).

It is quite possible that both groups of researchers are correct …that knowledge can be constructed by learners to solve problems, but that they need to be guided during the early stages of learning? This is the underlying idea of this dissertation, that instructors and instructional materials must guide learners during early schema acquisition. Later as learners become more expert, a constructivist philosophy can be useful, for it is then that they may be taught with ill-structured environments that require the construction of knowledge.
In 1993, Jonassen and his colleagues described Constructivist Learning Environments (CLEs) as being for more advanced learners, and expected more structured approaches for novices (Jonassen, Mayes, & McAleese, 1993).

However, Jonassen and his associates suggest the following:

“We believe that constructivistic learning environments may be used during the latter stages of knowledge acquisition and that they represent rich and meaningful environments for initial knowledge learners. However constructivistic environments are more reliably and consistently applied to support the advanced knowledge acquisition phase.” (Jonassen, Mayes, & McAleese, 1993, p.232)

Figure 4. Structured and ill-structured domains (adapted from Jonassen et al, 1993)

Jonassen et al. (1993) presented a continuum (Note Figure 4), suggesting that ill-structured domains be used later in the learning process. This figure even proposes that more structured learning domains like skill-based or procedure-based learning are
appropriately handled by well-structured instruction. This is an important point because this article shows that some of the most outspoken advocates for Constructivism believe constructivist learning environments have their limitations. Unfortunately the above argument seems to have been lost over the years, and now many educators believe constructivist learning environments are useful in all situations.

This then is the context of the controversy surrounding these issues. It is well understood that over the last twenty to thirty years that many educators have adopted a constructivist epistemology and that many may oppose this study from a philosophic or ideological basis. But the fundamental nature of a dissertation is to question the prevailing view, to generate new knowledge and new perspectives.

Finally, it should be stated that while the prevailing view is to support discovery or constructivism, good empirical studies have never lost favor with educational researchers. Thus before closing this discussion of discovery learning, it would be prudent to review a study by Sweller and his associates. This article is very important, because it actually compares learners who passively study worked examples with those using discovery-practice.

*Worked examples versus discovery learning*

Sweller (1988) suggests that discovering a problem solution constitutes a dual-task, requiring the learner to search for a problem solution, while trying to learn the underlying problem schema. Thus studying worked examples is a way to limit problem solving search, because worked examples only demonstrate the problem schema, eliminating irrelevant aspects of a problem solution (Sweller et al, 1998). However, given
this proposition cognitive load theory may be in conflict with discovery learning, one of America’s “most popular and frequently studied instructional procedures” (Paas, Renkl, & Sweller, 2004, p.6).

Tuovinen and Sweller (1999) compared learner performance given worked examples and discovery learning. In addition to studying these types of instruction they also compared the performance of those with and without experience. This experiment was carried out over three consecutive weeks. During the first week learners were asked to fill out and initial survey and then were introduced to FileMaker Pro (a Macintosh database program).

During week two learners were randomly assigned into two groups: an exploration group and a worked examples group. Both groups were given some basic instruction via a series of HyperCard lessons. The exploration group was given the following text-based instructions:

“Try out the functions in each of the lessons in situations you create-yourself, saving your files on the floppy disk provided. You may use any of the databases on the floppy disk if you wish. You will be asked to solve problems similar to the one shown in the lessons, in the test on this work. So direct your exploration towards gaining adequate mastery of the program to deal with such questions.” (Tuovinen & Sweller, 1999, p.337)

The worked example group was asked to read through a worked example which consisted of “a problem statement related to calculation or field construction or use and then an annotated step-by-step example of the way the problem could be solved with computer-screen views seen by the operator working to obtain the solution” (Tuovinen &
Sweller, 1999, p.337). These learners were subsequently asked to practice what they had learned on a similar problem.

Recall that worked examples have been only found to be useful with novices (the expertise reversal effect). To control for prior experience, levels of experience were measured in the initial survey. Learners were divided into two groups, those with database experience and those with no prior experience (true novices).

During the third week learners were tested with a common test composed of items similar to those taught in their lessons. Each learner was provided with a series of questions and required to create database files based on those questions. Test scores were analyzed with a 2x2 ANOVA (higher or lower levels of experience) X (worked-example or exploration group).

As expected the results showed a significant main effect with respect to levels of experience, with those with prior experience performing significantly better than those with less experience. But the main effect for groups was not significant. However there was a significant interaction between these variables (note figure 5).

When they compared the mean test scores for those without experience, they found those in the worked examples group performed significantly better than those in the exploration group (discovery practice). Thus, they again confirmed the worked example effect. As expected, they found means scores for the groups with prior experience were not significantly different, but that worked examples were not as beneficial for learners in this group (further evidence of the expertise reversal effect).
While these results are positive for cognitive load theorists, this was just one of many studies comparing learners who studied worked examples versus those who solved problems. Many, many, studies have confirmed these results showing overwhelming evidence in favor of worked examples as opposed to problems solving (Carroll, 1994; Paas, 1992; Paas & van Merriënboer, 1994; Quilici & Mayer, 1996; Sweller, 1988, Zhu & Simon, 1987). Finally according to Paas, Renkl, and Sweller (2004), there is comparatively little evidence demonstrating the efficacy of discovery learning.

Tuovinen and Sweller (1999) while confident of their results, closed their article with one caveat:
"It can, of course, be argued that exploration practice may be superior to worked examples, even for novices, if measures other than those of the present experiment are used. For example, exploration may favor long-term retention. Although this question must remain open until tested, it should be noted that in the present case, students with no previous database experience who learned by exploration, achieved such low test scores that minimal knowledge was available for long-term retention." (Tuovinen & Sweller, 1999, p.340)

Thus the current study intends to measure the cognitive load and learner performance one week after initial instruction. Given this is the case the next order of business is to discuss how that might be accomplished.

*How has cognitive load been measured?*

Our technology is just beginning to be able to peer inside the working brain to measure changes in brain function, thus it has been very difficult to measure cognitive load. But humans have been very imaginative and developed physiological and computational estimates of cognitive load. This section describes the various methods developed to date.

Brünken, Plass, and Leutner (2003) discuss the measurement of cognitive load in some detail, and even develop a classification scheme to describe cognitive load assessments. Brünken et al. (2003) classify cognitive load measurements along two basic dimensions: objectivity and causal relation. In their classification scheme, assessments along the causal relation dimension can be described as either direct or indirect, while those in the objectivity dimension are either objective or subjective (note Table 2).
Table 2. Methods for measuring cognitive load

<table>
<thead>
<tr>
<th>Objectivity</th>
<th>Causal Relationship</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Indirect</td>
</tr>
<tr>
<td>Subjective</td>
<td>Self-reported invested mental effort</td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td>Objective</td>
<td>Physiological measures</td>
</tr>
<tr>
<td></td>
<td>Behavioral measures</td>
</tr>
<tr>
<td></td>
<td>Learning outcome measures</td>
</tr>
</tbody>
</table>

Adapted from: Brünken, Plass, and Leutner (2003)

All of the measures employed by cognitive load researchers have their advantages and disadvantages. Tuovinen and Paas (2004) point out that most studies measuring cognitive load use the self-reported invested mental effort ratings developed in the early 1990s (note appendix D). However Brünken et al. (2003) suggest, even though Paas’ indirect, subjective, ratings scales seem useful, it still remains unclear how these mental effort ratings, are related to actual cognitive load.

On the other end of the spectrum, while functional magnetic resonance imaging (fMRI) provides direct objective cognitive load measurements, but this method is expensive and difficult for educational researchers to use with large populations. Nonetheless brain imaging is being used to study learners during computer-based problem solving (e.g. Anderson, Albert, & Fincham, 2005).

Brünken, Steinbacher, Plass, and Leutner (2002) is an example of a study that makes use of a dual-task assessment. They use a pre-test/post test design (pencil and
paper assessment) to measure learning given a dual task learning environment. Their materials included a split screen in which learners press a key given the color change of a stimulus. Dual-task methodologies are prominent in the psychological literature, and now a means of measuring cognitive load.

Given the constraints of this study, it is not possible to use fMRI, or regrettably dual-task measures like those of the Brünken et al. (2002) study. Instead this project will use a variety of techniques including performance-based measures and more indirect subjective measures like those proposed by Paas and van Merriënboer (1993). Thus, the next section explains these more indirect subjective measures.

*E is for efficiency*

In the early cognitive load literature, Sweller and others only used performance measures to measure cognitive load, then Paas (1992) began to propose that cognitive load was a “multidimensional concept” (p.429) (composed of perceived mental effort and mental load) and that “the intensity of [mental] effort is considered to be an index of cognitive load” (Paas, 1992, p 429). Paas and van Merriënboer (1993) further developed these ideas to produce a metric which they call “relative condition efficiency (E)” (Paas & van Merriënboer, 1993, p.737) which combines perceived mental effort and performance measures (e.g. performance time or accuracy).

To produce this construct, E (the relative condition efficiency measure), raw mental effort ratings and performance scores are standardized to provide a Z score for performance (P) and mental effort (R). From this set of Z scores, “E” is calculated (Paas and van Merriënboer, 1993) (note figure 6).
$$E = \frac{Z_{\text{performance}} - Z_{\text{mentaleffort}}}{\sqrt{2}}$$

Figure 6. Relative condition efficiency (E)

Figure 7 is a graph from the original Paas and van Merriënboer (1993) article. Condition scores above the base line where $E=0$, (in the upper left hand corner of the graph) are expected to have a greater relative condition efficiency. Thus these graphs show that the mean score for different instructional conditions can be compared.

Instructional conditions that are located above the line have a greater relative condition efficiency score because they have a higher group performance with lower invested mental effort (Paas & van Merriënboer, 1993). Conversely low instructional
efficiency is the result of low task performance and high effort (Tuovinen & Paas, 2004). Finally, it should be stated that this efficiency measure assumes mental effort and performance are linear with respect to one another (Paas & van Merriënboer, 1993).

The Paas and van Merriënboer (1993) study makes for a good example. It compared conventional problem solving (CONV), to worked examples (WORK), and completion problems (COMP). Paas and van Merriënboer (1993) later conducted a one-way ANOVA with these conditions to revealed a significant effect, $F(2,42) = 24.76$, $p < 0.001$. They conducted a post hoc procedure (Fischer’s test) to show that the conventional problem solving condition ($E=-1.15$) was significantly different from the other two conditions (worked example and problem completion group) which did not differ significantly.

It should be noted that Paas and van Merriënboer (1993) were quick to suggest that these scores should be qualified with performance data. Therefore even though the worked example and completion problem, relative condition efficiency scores did not differ significantly, that the mean performance scores for these groups are as follows: $\bar{x} = 78.57$ for the worked example group, $\bar{x} = 67.22$ for the problem completion group, and finally $\bar{x} = 51.60$ for the conventional discovery problem solving group. Thus it appears the worked example group out-performed both other groups, suggesting the worked example condition is the preferred, although not significantly different from the problem completion condition.
Paas et al (2003) found that researchers had used different scales to describe efficiency as a means of measuring cognitive load. It seems Paas and Van Merriënboer (1993) had used mental efforts estimates from the performance or test phase, while Sweller and his associates used mental effort estimates from the learning or acquisition phase (note Appendix D or table 3). To clarify matters Paas et al (2003) suggested dividing these into separate two efficiency metrics and that researchers use the terms “relative condition efficiency” (Paas and van Merriënboer, 1993, p.739) and “training or learning efficiency” (Paas et al, 2003, p.69).

Table 3. Relative condition efficiency and learning efficiency

<table>
<thead>
<tr>
<th>Acquisition phase</th>
<th>Performance phase</th>
</tr>
</thead>
<tbody>
<tr>
<td>Performance score</td>
<td>X</td>
</tr>
<tr>
<td>Mental effort estimate</td>
<td>X</td>
</tr>
</tbody>
</table>

*Table 3A* relative condition efficiency (adapted from Paas, 2007)

<table>
<thead>
<tr>
<th>Acquisition phase</th>
<th>Performance phase</th>
</tr>
</thead>
<tbody>
<tr>
<td>Performance score</td>
<td>X</td>
</tr>
<tr>
<td>Mental effort estimate</td>
<td>X</td>
</tr>
</tbody>
</table>

*Table 3B* learning efficiency (adapted from Paas, 2007)

Each of these metrics signifies different aspects of the learning-testing process (Tuovinen & Paas, 2004). Paas et al (2003) suggests *relative condition efficiency* is mostly concerned with mental effort expended during a test performance, and may be more related to transfer. The main benefit of this metric is that it helps instructional designers to measure the efficiency of instructional conditions relative to a test performance (Tuovinen and Paas, 2004).
Learning efficiency is the metric first introduced by Sweller and his associates in 1996, and is more concerned with mental expenditures during learning rather than during a test or performance (Marcus, Cooper, and Sweller, 1996). This is important because this metric provides an immediate test of the efficiency of the learning environment as opposed to the test environment.

3 dimensional approaches

Appendix D shows most studies since the original Paas and van Merriënboer (1993) article use learning efficiency rather than their relative condition efficiency. Even though this is the case, Touvenin and Paas (2004) defend the original metrics because they say it is important to measure the relative importance of mental effort during a test or performance:

“It is quite feasible for two people to receive the same performance scores, while one of them needs to work laboriously through a very effortful process to arrive at the same number of correct answers, whereas the other person reaches the same answers with a minimum of effort.” (Tuovinen & Paas, 2004, p.140).

Thus Tuovinen and Paas began to consider new metrics that would consider both the test phase and the learning phase simultaneously. In their search for this new metric Tuovinen and Paas (2004) created some new terminology. They refer to both relative condition efficiency and learning efficiency as 2 dimensional (2D) measures because they include one mental effort measure and one performance measure.
Tuovinen and Paas (2004) also develop a 3D (or 3 dimensional measure) by combining the factors of the 2D measures; to do so they factor in learning effort (E_L), test effort (E_T), and performance (P) in the following formula:

\[
3D \text{ Efficiency} = \frac{P - E_L - E_T}{\sqrt{3}}
\]

Figure 8. Overall or 3D efficiency

Like 2D efficiency metrics, this 3D measure is also graphed, but in three dimensional space. Tuovinen and Paas (2004) claim this combines the best features of both metrics. However Salden, Paas, Broers, and van Merriënboer (2004) produced yet another 3D metric that makes use of total training time. To do so Salden et al (2004) combine performance (P), mental effort (ME) and total training time (TT) in the following formula:

\[
\text{Training Efficiency} = \frac{P - ME - TT}{\sqrt{3}}
\]

Figure 9. Training efficiency

This metric makes more sense because it factors in time. Certainly cognitive load is concerned with the amount of time involved during training. This metric can also be graphed in 3 dimensional space with performance, mental effort and time on the three axes. Finally other studies since Tuovinen and Paas’ article have begun to use 4D metrics which include motivation (Nadolski, Kirschner and van Merriënboer, 2005). However 4D metrics will not be used in this study. Certainly motivation and mental effort are
interesting, but are they necessary for an efficiency equation? Or does this include unnecessary subjective elements?

Recall that Brünken, Plass, and Leutner (2003) separate cognitive load measures based upon direct and indirect, or by subjective and objective methods. This dissertation study intends to derive a new formula based completely on the objective measures of task performance (P) and performance time (T), but does not include the subjective mental effort ratings. This study describes this as performance efficiency:

\[ \text{Performance Efficiency} = \frac{P - T}{\sqrt{2}} \]

*Figure 10. Performance efficiency*

While this new metric does not represent mental efficiency, it is an objective efficiency metric that can be used to describe the learner’s performance given the cognitive load of the instructional condition. Before Paas and van Merriënboer (1993) article, performance measures were used exclusively. This is just a simpler way to express learner performance given time.

Performance in most of the cognitive load studies, as with most instructional research is measured via paper and pencil tests (e.g. Paas and van Merriënboer, 1993; Sweller, 1988, Brünken, Plass, and Leutner, 2004). Rather than using paper and pencil tests like the Brünken et al (2002) study, the current study intends to use a more direct objective performance measure.

In summary, there is little doubt that this Paas and van Merriënboer’s efficiency metric has had a dramatic effect on the cognitive load literature. It has helped a number
of researchers to produce an estimate of cognitive load. The current study intends to use the relative condition efficiency metric described by Paas and van Merriënboer (1993), the learning efficiency metric described by Paas et al (2003) (note table 3), and finally the simplified objective performance efficiency described above (figure 10). The details of how these metrics will be used in this study are explained in chapter 3.

Rationale

The general hypothesis of this project is that animated demonstrations function like worked examples. If this is the case several indicators would be apparent given a comparison of the right combination of instructional conditions. This section outlines the reasoning behind the instructional conditions used in this study.

Given the hypothesis of this dissertation, it is suggested that if learners are provided with animated demonstrations that these learners will demonstrate increased performance over their problem solving peers (the worked example effect). In addition, Tuovinen and Sweller (1999) put forward the notion that retention may not be as durable with worked examples as it is with discovery problem solving. Thus it could be argued that Palmier’s animation deficit is a legitimate concern given animated worked examples. Therefore the primary question of this dissertation becomes: is this retention deficit a concern given animated demonstrations over longer retention intervals? Thus a comparison of animated demonstrations to discovery-based problem solving is suggested.

Paas and van Merriënboer (1994) also had an important suggestion for instructional designers. They suggest learners should be provided with a variety of worked examples (Paas & van Merriënboer, 1994). Thus the question becomes if one provides learners with different animated demonstrations will that promote schema
acquisition (the variability effect)? So the effects of different animated demonstration should be compared. Finally cognitive load is measured with objective or subjective measures (Brünken et al, 2003). This study intends to use a combination of both objective performance measures and subjective mental effort ratings to estimate cognitive load by measuring the mental efficiency of instructional conditions. Thus there are several outstanding research questions:

**Question 1:** Will learners using one of the instructional conditions complete tasks in significantly less time than learners in the other conditions? (Performance time)

**Question 2:** Will learners using one of the instructional conditions be significantly more accurate? (Accuracy)

**Question 3:** Will one of the instructional conditions be significantly more efficient during initial training? (Learning efficiency)

**Question 4:** Will one of the instructional conditions be significantly more efficient during delayed test? (Instructional efficiency)

**Question 5:** Will the group performance (from one of the instructional conditions) be significantly more efficient during delayed test? (Performance efficiency)

Questions 1 and 2 will answer the general hypothesis of this study: if animated demonstrations act like animated worked examples (the worked example and variability effect). Questions 3, 4, and 5 are concerned with the efficiency of the instructional conditions being compared. Chapter 3 explains how the current study intends to answer these research questions. Specifically it outlines the data collection techniques and outlines the statistical procedures used to analyze these questions.
CHAPTER 3 - METHODS

This dissertation project questions if animated demonstrations act as animated worked examples. To do so, it tests the worked example effect and the variability effect, an extension of the worked example effect, using animated demonstrations. Specifically, it contrasts the cognitive load and learner performance of those using animated demonstrations with those using discovery-based problem solving. This chapter outlines the methodology of this study.

Research Design

Given there are multiple outcome variables (performance time and accuracy) a multivariate data analysis is required. A multivariate analysis of variance (MANOVA) will be used as the statistical procedure to analyze performance time and accuracy. In addition relative condition and learning efficiency will be analyzed with the procedure originally developed by Paas and van Merriënboer (1993) and Paas, Tuovinen, Tabbers & Van Gerven (2003) respectively. In addition a variant of these efficiency measure “performance efficiency” will also be measured.

Sample

The sample used in this study will be pre-service teachers. Participants in this study will be undergraduates of EME 2040, an introductory instructional technology course at a large southeastern university (the University of South Florida). More importantly, subjects will be randomly assigned to four groups, making this an experimental study. These learners have been chosen as a target population because they are novices.
Finally an a priori power analysis for a four group MANOVA produces a sample size of n=115 participants. This number of participants is necessary to arrive at a power of .80, with a small effect size $q^2= 0.125$, given an $\alpha$ of .05 (Stevens, 2002) thus participants will sampled across two semesters.

**Research questions**

The following research questions have been determined from the literature review:

**Question 1:** Will learners using one of the instructional conditions complete tasks in significantly less time than learners in the other conditions? (Performance time)

**Question 2:** Will learners using one of the instructional conditions be significantly more accurate? (Accuracy)

**Question 3:** Will one of the instructional conditions be significantly more efficient during training? (Learning efficiency)

**Question 4:** Will one of the instructional conditions be significantly more efficient during the delayed test? (Instructional condition efficiency)

**Question 5:** Will the group performance from one of the instructional conditions be significantly more efficient during the delayed test? (Performance efficiency)

**Expectations**

Sweller and Cooper (1985) found that learners who studied worked examples took significantly less time to solve problems with fewer errors; this is precedence for the current project. Because of Sweller and Cooper’s initial findings, it is expected that learners who study animated demonstration (animated worked examples) will take less time to solve problems [performance time (S)] be more accurate [accuracy (A)] than learners who...
practice problem solutions. Given there are multiple outcome variables research questions one and two will be answered with a MANOVA (note table 4).

*Table 4. Research questions one and two described as hypotheses*

\[
\begin{bmatrix}
\mu_{1S} \\
\mu_{2A} \\
\mu_{3A}
\end{bmatrix} = \begin{bmatrix}
\mu_{1A} \\
\mu_{2A} \\
\mu_{3A}
\end{bmatrix} \quad H_0 = \text{There is not a significant difference in performance given the type of instruction}
\]

\[
\begin{bmatrix}
\mu_{1S} \\
\mu_{2S} \\
\mu_{3S}
\end{bmatrix} \quad H_a = \text{There is a significant difference in performance given the type of instruction}
\]

*Materials*

*Introductory survey*

Given the expertise reversal effect (Kalyuga, Ayres, Chandler & Sweller, 2003), it is imperative that this study recruit participants who have little or no prior knowledge of the subject matter. Learners with more expertise will be treated in a manner similar to the learners in the Tuovinen and Sweller (1999) study.

Therefore, it will be necessary to develop a pretreatment survey to determine the number of participants who have prior knowledge with the subject matter (Adobe Photoshop editing techniques). This will be a web-based survey developed with Microsoft FrontPage® 2003 (Microsoft, 2003a) and administered prior to the instructional treatment. This questionnaire will also obtain basic demographic and learner expertise questions (note Appendix F).
Overview

This study contrasts learner performance given four instructional treatments (note table 5). The Paas (1992) and Tuovinen and Sweller (1999) studies both provided their learners with an overview before administering the instructional conditions. Thus all treatments will be preceded by a brief overview giving learners some context. This overview will be a short web-based presentation (~ 2 minutes) developed with Techsmith Camtasia 4.0 (Techsmith, 2006) and will provide learners with a brief introduction to graphic design and digital image manipulation. This presentation presents learners with screenshots of Adobe Photoshop Elements 2.0 and describes several interface elements (Adobe Systems, 2002). Once this overview concludes a JavaScript separates learners into four separate instructional conditions.

Instructional Materials

The four instructional conditions are: an animated demonstration (demo only), an animated demonstrations with the Mr. Potatohead problem (demo+ practice), a different animated demonstration, with the Mr. Potatohead problem (demo2+practice), and finally a discovery practice condition with the Mr. Potatohead problem (discovery-practice or practice only condition).

Both animated demonstrations were developed with Techsmith Camtasia 4.0 (Techsmith, 2006). These materials teach a novice how to use Photoshop layers. Specifically learners will learn how to select, move, rotate, and hide layers within an Adobe Elements document (Adobe Systems, 2002).
### Table 5. Instructional materials

<table>
<thead>
<tr>
<th>Condition 1</th>
<th>Condition 2</th>
<th>Condition 3</th>
<th>Condition 4</th>
</tr>
</thead>
<tbody>
<tr>
<td>Demo Only</td>
<td>Demo + Practice</td>
<td>Demo 2+ Practice</td>
<td>practice only</td>
</tr>
</tbody>
</table>

- **Condition 1 (demo only)** – The instructional component will be a brief animated demonstration, that only shows the learner how to put together a Mr. Potatohead document with Photoshop. This tutorial only demonstrates a series of Photoshop procedures.

- **Condition 2 (demo + practice)** (~10 minutes long) – This condition will differ from condition 1, in that these learners will view the same demonstration as those in condition 1, but also use Adobe Photoshop Elements 2.0 to put together the Mr. Potatohead problem.

- **Condition 3 (demo 2+ practice)** (~10 minutes long) - Learners will watch a different animated demonstration which demonstrates the same underlying
skills as demo 1, but puts together a photo collage instead. After watching this demonstration learners will put together the Mr. Potatohead problem.

- Condition 4 (practice only) – Learners in this condition will receive no additional instruction other than the overview, but will be asked to put together the Mr. Potatohead problem.

Condition 2, 3, and 4, will attempt the Mr. Potatohead problem (Note table 6). This is a disassembled cartoon figure that requires reassembly within Adobe Photoshop Elements 2.0. Specifically learners must move, rotate, flip and reorder the parts of this cartoon figure (Photoshop layers) to reassemble the character.

Table 6. Week one- the Mr. Potatohead problem

<table>
<thead>
<tr>
<th>Disassembled Mr. Potatohead</th>
<th>Reassembled Mr. Potatohead</th>
</tr>
</thead>
<tbody>
<tr>
<td><img src="image1" alt="Disassembled Mr. Potatohead" /></td>
<td><img src="image2" alt="Reassembled Mr. Potatohead" /></td>
</tr>
</tbody>
</table>

The Picnic Problem

One of the goals of this project is to understand the long term influences of animated demonstrations, so one week after initial instruction (week two) learners will put together another Adobe Photoshop Elements document (the Picnic problem, note table 7).
Table 7. Week two - the picnic problem

<table>
<thead>
<tr>
<th>Disassembled Picnic</th>
<th>Reassembled Picnic</th>
</tr>
</thead>
</table>

This new document requires the same skills as the Mr. Potatohead problem, but their newly learned skills must be used in a new context. The picnic problem is a construction task that requires learners to put together two stick figures within a scene. The picnic problem is somewhat more complicated than the week one practice problem (the Mr Potatohead document) because it is composed of multiple figures within a scene, along with several other objects (a picnic basket, picnic table, an umbrella, and birds).

Procedure

This section of the proposal outlines the procedures for data collection and measurement of the dependent variables (Appendices A, B and C describing these procedures). Appendix A is a flowchart that summarizes how the facilitator should interact with participants during data collection. Appendices B and C describe the use of Morae Manager as a means of collecting performance time and learner efficiency data.
Data collection

The learning environment

A facilitator must work with university technical support to insure that several programs are installed on the computer stations used for this study. This study requires a web browser, the Adobe Flash plug-in (7.0 or greater), Adobe Photoshop Elements 2.0, and Techsmith Morae Recorder.

A facilitator must prepare the environment before learners enter the room. First this facilitator must confirm that all necessary software for the study has been installed. Therefore it is best if the facilitator is accompanied by technical support personnel, so they may alter software installations if necessary. Once each computer has been checked for the appropriate software, a set of earphones must be plugged into each system. Next, the facilitator should check volume levels on all computers. Finally, shortly before data collection is to begin, the facilitator must go to each computer station to start Morae Recorder. Finally a folder full of short-cuts and the day’s Photoshop problem must be placed on the desktops of all computers.

Acquisition phase

Once learners enter the learning environment, it will be explained that they are being asked to volunteer to participate in a research study. They will be asked to sit at an appropriate computer (one with earphones). Only those stations with the required software will have earphones plugged in. Participants will be asked to move if they sit at an inappropriate station. Once learners are seated, learners will be handed the Institutional
research board (IRB) documentation and be asked to read and sign it. Learners will also be asked to print their name on an index card which has the computer number written on it.

Once this paperwork has been signed, learners will be instructed to put their earphones on. These will provide learners with an individualized learning experience. In addition each participant will need to wear earphones to insure that they do not hear instruction from other computers. Even though condition 4 will not receive instruction other than the overview, their earphones will protect other participants from hearing computer audio feedback.

Once the above conditions have been met, the acquisition phase will begin. This will begin with a scripted introduction. This introduction will explain to learners that:

- they are taking part in a dissertation study;
- that this study will be conducted over a two week period (the acquisition phase, and delayed test);
- that at some point they may be required to use the computer to work through a problem scenario (the Mr. Potatohead puzzle during week one);
- that they can not be helped to put together the puzzle that they will have to figure it out on their own;
- and finally that their onscreen behavior will be recorded.

All learners will be told to open a folder on the desktop of their computers, and to double-click on an icon. This icon is a shortcut to a web-based survey (Note Appendix F). This survey will welcome the learner to the study and provide them with a brief demographic survey. Once they answer all questions and click submit, a JavaScript will
randomly assign the learner to an instructional condition. Next learners will take part in one of these instructional conditions which may include an animated demonstration and/or the reassembly of the Mr Potahead puzzle (note table 5).

After taking part in the instructional condition, learners in Condition 1 (demo only group) will be asked to fill out the week one survey #2 (Appendix G). Learners in Conditions 2, 3, and 4 will be asked to launch the Mr. Potatohead problem. Learners will not be given help if they request it, however they will be politely told they can quit if they feel they cannot finish the problem on their own.

All groups of learners will conclude week one activities by completing the week 1 survey #2. Learners will be asked to leave the environment when they are finished with the survey. As learners leave the room they will be thanked for their participation and instructed not to study or use Adobe Photoshop, or to discuss the instructional materials for the next week.

Once all learners have left the room the primary researcher will walk to each station and save their files for later analysis. These files will be saved according to the computer number (e.g. “7.rdg” rdg is the Morae file 3 letter extension). Recall that the computer number is also written on the index card that the learners printed their names on.

**Performance Phase**

The delayed test phase will be conducted one week after initial instruction. It is during this delayed test (with the picnic problem) that data will be collected concerning the dependent variables: performance time, accuracy and learner efficiency.

The learning environment will have to be prepared much as during the previous week, insuring that all week one files have been deleted and replaced with week two files.
In addition Morae Recorder must be turned on. Next learners will be let into the environment and asked to complete the picnic problem. After learners have completed the picnic problem, they will be asked to complete the week two survey (note Appendix H). If they decide that they cannot finish the problem then they will be asked to complete the week two survey. Once the survey is complete, they will be asked to leave and be thanked for their participation. Their recorded performance will be saved in a manner somewhat similar to the initial test files (e.g. “7.rdg”) but kept in a separate directory from the week one files.

Dependent variables

Gagné (1964) discusses two general categories of dependent variables associated with problem solving studies. He suggests most researchers are concerned with 1) “the rate of attainment of some criterion performance”, (performance time) and 2) “the degree of correctness of this performance” (learner efficiency) (Gagné, 1964, p.295).

Performance Time

Gagné’s “rate of attainment” is easily measured as the performance time in seconds. For the duration of this study, this variable may also be described as speed, in keeping with Palmiter’s (1991) nomenclature. As stated in chapter 1, learner performance (performance time) will be recorded and coded within TechSmith Morae® 1.01 (Techsmith, 2004) (note Appendix B).

Accuracy

Gagné’s second category, “the degree of correctness” is not as easily defined. Gagné (1964) mentions single problem solutions are usually measured as either pass or fail.
In other words, either the learner correctly solves the problem (attains the problem goal) or not. While problem completion is an important part of any problem solving study, this pass/fail measurement oversimplifies the learner’s attempt during problem solving.

Gagné also mentions that one may score a learner’s performance as “partially correct” (Gagné, 1964), in which case, a partial solution suggests that some learning has occurred. This dependent variable will be described as accuracy in keeping with Palmiter’s (1991) nomenclature. Accuracy will be measured using a rubric based upon the Picnic problem (note Appendix I). This rubric is based upon the number of problem solving operators required to solve the problem.

Relative condition and learning efficiency

The reader may recall from chapter 2 that both “relative condition efficiency” and “learning efficiency” was described by Paas, Tuovinen, Tabbers & Van Gerven (2003). Both instructional condition efficiency and learning efficiency are a combination of a performance metric and perceived mental effort. The difference between the two is when the perceived mental effort score is recorded (note table 8). Both will be analyzed in this project.

Measurement Procedures

As described in Chapter 1, TechSmith Morae is the primary tool for data collection (Techsmith, 2004). Certainly this method is not flawless, but given researchers follow the same procedure they can produce identical results (note Appendix B and C for job aids of these procedures). The picnic problem (week two) will be the only performance analyzed.
Procedure for measuring performance time

The “overall performance time” will begin when the learner first moves the cursor, after the picnic document has been opened. At this point a researcher will code that point on the timeline as an “in point.” The researcher will watch what the learner does during their attempt. When the learner has ceased to move the cursor, the researcher may code that point on the timeline as the “out point.” Given that an “in point” and “out point” have been set, these two positions may be compared to produce what Morae describes as a “segment.” The “overall performance time” then is a segment of video between the “in point” and “out point.” Morae analyzer displays the duration of segments in seconds, the performance time, which may be logged for later analysis.

Procedure for measuring accuracy

Researchers will view the recorded video files of a learner’s on screen action and score that file with a rubric developed for this purpose (please note Appendix I). Behavioral analysis data forms are generally organized into a tabular format (Hinde, 1973; Lehner, 1996), similar to those found within a spreadsheet program like Microsoft Excel® (Microsoft, 2003b). Excel documents will be used for the duration of this project for ease of use and storage.

An accuracy rubric (a Microsoft Excel document) will be used to score a learners attempt, by granting 1 point for moving or rotating a layer into the correct position within the scene. In addition because the main objective of this tutorial is to teach the learner how to manipulate Photoshop layers, 2 points will be granted for raising or lowering a layer correctly (relative to the other layers surrounding it) and finally an additional 2 points will be granted for flipping a layer horizontal.
Procedure for measuring relative condition efficiency

Paas and van Merriënboer (1993) described a metric for measuring the relative efficiency of instructional conditions, based upon a combination of performance scores and mental effort ratings. To produce the mental effort ratings learners in the current study will fill out a survey question (a 9-point mental effort rating) following their performance with the picnic problem. This survey question is identical to the one use in the Paas and van Merriënboer (1993) study (note Appendix H).

Paas and van Merriënboer (1993) used the test scores from a statistics test (originally reported in Paas, 1992) with the percentage correct as their “raw score” (p.429). Since the current study uses a single performance problem, the raw performance score for this study will be the score obtained from the accuracy rubric (this procedure is described above). Thus the current study will follow the Paas and van Merriënboer (1993) procedure by standardizing the mental effort ratings and the accuracy scores from the week two performance, to produce a performance z scores for each individual.

Paas and van Merriënboer (1993) used group mean scores to replace missing scores. This rule will also be followed in this study. Once a list of z scores is developed, E (the relative condition efficiency) can be computed from the following formula for each individual:

$$E = \frac{Z_{\text{performance}} - Z_{\text{mentaleffort}}}{\sqrt{2}}$$

Finally each instructional condition will receive its own group mean efficiency score, which will be graphed in a manner to those similar to figure 7 (in chapter 2).
Relative condition efficiency is then calculated to be the perpendicular distance to the efficiency equals zero line (Paas & van Merriënboer, 1993). According to Paas and van Merriënboer, (1993)

“The sign of the relative condition efficiency is dependent on $R$ and $P$ according to the following rule: If $R - P < 0$, then $E$ was assigned a positive value and if $R - P < 0$, then $E$ was assigned a negative value” (Paas & van Merriënboer, 1993, p.742).

Like the Paas and van Merriënboer (1993) study, the current study will conduct a one-way ANOVA with the group means. Post hoc comparisons will be used to determine if they are significantly different.

Table 8. Relative condition efficiency and learning efficiency

<table>
<thead>
<tr>
<th>Acquisition phase</th>
<th>Performance phase</th>
</tr>
</thead>
<tbody>
<tr>
<td>Performance score</td>
<td>X</td>
</tr>
<tr>
<td>Mental effort estimate</td>
<td>X</td>
</tr>
</tbody>
</table>

Table 8A) relative conditional efficiency (adapted from Paas, 2007)

<table>
<thead>
<tr>
<th>Acquisition phase</th>
<th>Performance phase</th>
</tr>
</thead>
<tbody>
<tr>
<td>Performance score</td>
<td>X</td>
</tr>
<tr>
<td>Mental effort estimate</td>
<td>X</td>
</tr>
</tbody>
</table>

Table 8B) learning efficiency (adapted from Paas, 2007)

Procedure for measuring learning efficiency

Paas et al (2003) describe the metric used by Sweller and his associates as learning efficiency; this is the combination of perceived mental effort ratings during instruction and the subsequent test performance scores (Paas et al, 2003). The procedure for measuring learning efficiency is very similar to that described for relative instructional condition
efficiency. But rather than using the mental effort ratings from week two (relative instructional condition efficiency) learning efficiency uses the week one mental effort ratings (note table 8).

Mental effort and performance scores are standardized to provide a Z score for performance (P) and mental effort (R). From this set of Z scores, “E” (in this case, learning efficiency) is calculated using the formula in Figure 11 (Paas and van Merriënboer, 1993).

\[
E = \frac{Z_{\text{performance}} - Z_{\text{mentaleffort}}}{\sqrt{2}}
\]

Figure 11. Paas and van Merriënboer’s efficiency formula

In order to obtain mental effort ratings during the acquisition phase (week one) learners will fill out a survey with a mental effort question identical to the one used in the Paas and van Merriënboer (1993) study (note Appendix G). Finally a one-way ANOVA will be used with the group means, followed by post hoc comparisons to determine if they are significantly different.

Procedure for measuring performance efficiency

The current study intends to produce a simplified version of the above efficiency measures. This measure called performance efficiency only relies on the objective measures of performance (P) and total performance time (T) and does not include a mental effort rating:
Performance Efficiency $= \frac{P - T}{\sqrt{2}}$

The performance score in this study will be the score provided by the accuracy rubric (Appendix I). This metric will be graphed like the other efficiency metrics. In addition a one-way ANOVA will be used with the group means, followed by post hoc comparisons to determine if they are significantly different.

Reliability

Summer and fall participants

Because the power analysis suggested a sample size of 115 participants, it will be necessary to collect data across two semesters, given the size and offerings of the EME 2040 classes. Unfortunately several months will pass between semesters and since the learners will be sampled over the summer and fall semesters and from two different classrooms (summer in EDU213 and fall in EDU417) the data from these groups may differ. Thus it will be necessary to see if these potential differences will influence the data set. A Hotelling’s $T^2$ will be run to see if these groups differ significantly with respect to performance time and accuracy.

Inter-observer reliability

Observational data has its advantages and disadvantages. While it may be a more direct method of observing behavior, with less conceptual interference from tests or questionnaires, this type of data has its own issues, like coding errors and observer drift (Talpin & Reid, 1973; Knupfer & McLellan, 1996). To avoid these potential errors this study will check the reliability of the data by using inter-observer reliability estimates.
Mook defines inter-observer reliability as “the degree to which independent observers observe the same result” (Mook, 2007).

In the current study, the independent observers will be one individual, making independent observations at different times. The observations can be compared by using the following formula (CCNY PSY, 2007):

\[
\text{Interobserver Reliability} = \frac{\# \text{ of agreements}}{\# \text{ of opportunities to agree}} \times 100
\]

These estimates will be made for a randomly selected group of participant data files (n=40). Finally the inter-observer reliability estimates will only be conducted on participant data files with the performance time and accuracy procedures.

Thank you for reviewing this proposed study.
REFERENCES


89


Moore, M.G. (1989). Three types of interaction. American Journal of Distance Education. 3 (2) 1-6.


Techsmith (2006). *TechSmith Camtasia Studio 4.0* [Computer program] Okemos, MI

Techsmith (2004). *TechSmith Morae 1.0.1* [Computer program] Okemos, MI


APPENDIX A: INSTRUCTIONAL PROCEDURES DURING WEEK ONE

Facilitator Installs Morae recorder

Facilitator turns on recorder

Learners enter test environment

Facilitator reads a carefully prepared script

Learners open and answers survey 1

Learners are randomly assigned into a group

Learner interacts with Condition 1
Learner interacts with Condition 2
Learner interacts with Condition 3
Condition 4 learner instructed to continue

Learner opens and solves Mr. Potatohead document

Learner opens and completes Week 1 survey #2

Learner leaves test environment

Facilitator turns off Morae Recorder

Facilitator saves learner file to flash drive
APPENDIX B: MEASURING PERFORMANCE TIME

TechSmith Morae

PROCEDURE FOR MEASURING PERFORMANCE TIME (SPEED)

1. Click the search button
2. Select Keyboard, Mouse Clicks, Web Page Changes, Markers, and Segments.

3. Click “Search Now”

4. Select a mouse click (e.g. when a learner selects the eyes)

5. Click the “set In point” button
6. click the set marker button (dialog box opens)

7. Label the marker as shown to the right (e.g. "begin move eyes")

8. Click OK

9. select an event when the learner finishes the procedure (e.g. set the final position of the eyes)
10. Click the “set out point” button

11. Click the set marker button (dialog box opens)

12. Label the marker as shown to the right (e.g. “end move eyes”)

13. Click OK

14. Click the “set end point” button
15. click create segment button (dialog box opens)

16. label segment (e.g. "end move eyes")

17. click ok
APPENDIX C: MEASURING LEARNER EFFICIENCY

1. Click Search

2. If “Windows” is checked, uncheck it

3. Check Mouse Clicks
4. Click In and out points

5. Log Total
APPENDIX D: MENTAL EFFORT AND PERFORMANCE BY PHASE

<table>
<thead>
<tr>
<th>Studies</th>
<th>Acquisition phase</th>
<th>Performance phase</th>
</tr>
</thead>
<tbody>
<tr>
<td>Paas and Van Merrienboer (1993)</td>
<td></td>
<td>ME, P</td>
</tr>
<tr>
<td>Marcus, Cooper and Sweller (1996)</td>
<td>ME P</td>
<td></td>
</tr>
<tr>
<td>Tindall-Ford, Chandler, and Sweller (1997)</td>
<td>ME P</td>
<td></td>
</tr>
<tr>
<td>Yeung, Jin, and Sweller (1997)</td>
<td>ME P</td>
<td></td>
</tr>
<tr>
<td>Kalyuga, Chandler, and Sweller (1998)</td>
<td>ME P</td>
<td></td>
</tr>
<tr>
<td>Kalyuga, Chandler, and Sweller (1999)</td>
<td>ME P</td>
<td></td>
</tr>
<tr>
<td>Tuovinen and Sweller (1999)</td>
<td>ME P</td>
<td></td>
</tr>
<tr>
<td>Yeung (1999)</td>
<td>ME P</td>
<td></td>
</tr>
<tr>
<td>Kalyuga, Chandler, and Sweller (2000)</td>
<td>ME P</td>
<td></td>
</tr>
<tr>
<td>Camp, Paas, Rikers, and Van Merrienboer (2001)</td>
<td>ME, P</td>
<td></td>
</tr>
<tr>
<td>Kalyuga, Chandler, and Sweller (2001)</td>
<td>ME P</td>
<td></td>
</tr>
<tr>
<td>Kalyuga, Chandler, Tuovinen, and Sweller (2001)</td>
<td>ME P</td>
<td></td>
</tr>
<tr>
<td>Van Merrienboer, Schuurman, De Croock, and Paas (2002)</td>
<td>ME P</td>
<td></td>
</tr>
</tbody>
</table>

Note. Adapted from Tuovinen and Paas (2004) ME = mental effort, P = performance.
## APPENDIX E: ANIMATED DEMONSTRATION STUDIES

<table>
<thead>
<tr>
<th>Authors</th>
<th>Subject matter</th>
<th>Instructional conditions</th>
<th>Results</th>
</tr>
</thead>
<tbody>
<tr>
<td>Palmiter and Elkerton</td>
<td>HyperCard</td>
<td>- Text only,</td>
<td><strong>Speed (Performance time):</strong> a significant session x media interaction ($F_{2,42}=7.06, p&lt;0.003$) with both demonstration groups completing tasks in significantly less time than text group at the initial test.</td>
</tr>
<tr>
<td>(1991)</td>
<td>12 procedures</td>
<td>- Animated demonstration,</td>
<td><strong>Accuracy:</strong> There was a significant session x media interaction ($F_{2,42} = 9.97, p &lt; 0.001$) with the demonstration groups completing significantly more correct trials than the text group at the initial test.</td>
</tr>
<tr>
<td>(n=48)</td>
<td></td>
<td>- Animated demonstration w/text</td>
<td><strong>Retention:</strong> There was no significant difference between the groups in performance time a week later.</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Immediate test, &amp; delayed test</td>
<td><strong>Transfer:</strong> The demonstration groups completed similar task in less time during the initial session than the text group. However a significant Session x Media interaction ($F_{2,42} = 3.64, p &lt; 0.04$) was found in which the there was a significant increase in time for the demonstration groups between the training and delay sessions; although there was a increase in the time required by the text group between sessions it was not significant.</td>
</tr>
<tr>
<td>Study</td>
<td>Software</td>
<td>Procedures</td>
<td>Immediate Test</td>
</tr>
<tr>
<td>--------------------------------------</td>
<td>-------------------</td>
<td>------------</td>
<td>----------------</td>
</tr>
<tr>
<td>Waterson &amp; O’Malley (1993)</td>
<td>Cricket Graph</td>
<td>6 procedures</td>
<td></td>
</tr>
<tr>
<td>(n=30)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Lipps Trafton and Gray (1998)</td>
<td>Microsoft Excel</td>
<td>12 procedure</td>
<td></td>
</tr>
<tr>
<td>(n=64)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Performance time:</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>The combination group completed tasks in significantly less time than text group given identical tasks ($F_{2,54}=14.08$, $p&lt;0.01$), and similar tasks ($F_{2,54}=9.85$, $p&lt;0.01$), but not significantly different given different tasks ($p=0.07$)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Accuracy:</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>($F_{1,60}=9.56$, $p &lt; .005$, MSE =.01) demonstration group significantly more accurately at the than text group acquisition session</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Performance time:</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Animation group performed tasks in significantly less time than text group (but not at delayed test)</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
APPENDIX F: INITIAL DEMOGRAPHICS SURVEY

Survey

Thank-you for agreeing to be a part of this study. Please answer the following questions:

What is your gender?
○ male ○ female

What is your academic status?
○ Freshman ○ Sophomore ○ Junior Senior
○ Post Graduate ○ Graduate ○ Non degree seeking

How old are you?
○ 18 ○ 19 ○ 20 ○ 21 ○ 22 ○ 23 ○ 24 ○ 25
○ 26 ○ 27 ○ 28 ○ 29 ○ 30
○ 31-35 ○ 36-40
○ 41-49 ○ 50 or older

Is English your first language?
○ yes ○ no

Have you scanned a document before?
○ yes ○ no ○ I don't know

Have you edited a digital image before?
○ yes ○ no ○ I don't know

Have you used Adobe Photoshop (or Adobe Photoshop Elements)?
○ yes ○ no ○ I don't know

Please indicate how often you use programs like Adobe Photoshop (or Adobe Photoshop Elements)?
○ never ○ seldom ○ sometimes ○ often ○ very often
APPENDIX G: WEEK ONE SURVEY #2

Survey 2

Earlier you participated in an instructional activity...

1. How would you describe last week’s activities… I invested:
   ○ very, very low mental effort
   ○ very low mental effort
   ○ low mental effort
   ○ rather low mental effort
   ○ neither low nor high mental effort
   ○ rather high mental effort
   ○ high mental effort
   ○ very high mental effort
   ○ very, very high mental effort

2. Did you feel the instructional materials were useful?
   ○ Strongly agree
   ○ Agree
   ○ Not sure
   ○ Disagree
   ○ Strongly disagree

3. How would you improve the instruction
APPENDIX H: WEEK TWO SURVEY

Week 2 Survey

Earlier today you participated in an instructional activity...

1. How would you describe this activity...

   I invested:
   ○ very, very low mental effort
   ○ very low mental effort
   ○ low mental effort
   ○ rather low mental effort
   ○ neither low nor high mental effort
   ○ rather high mental effort
   ○ high mental effort
   ○ very high mental effort
   ○ very, very high mental effort

2. Did you use Adobe Photoshop Elements (or Adobe Photoshop) since we met last week?

   ○ Yes
   ○ No
## APPENDIX I: ACCURACY RUBRIC

<table>
<thead>
<tr>
<th>rotate</th>
<th>layer</th>
<th>move</th>
<th>Points assigned</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>***</td>
<td>***</td>
<td>鸟1 – move</td>
<td>1</td>
<td>move to location near other birds in the sky (left hand side of screen)</td>
</tr>
<tr>
<td>***</td>
<td>***</td>
<td>鸟2 – move</td>
<td>1</td>
<td>move to location near other birds in the sky (left hand side of screen)</td>
</tr>
<tr>
<td>***</td>
<td>***</td>
<td>鸟3 – move</td>
<td>1</td>
<td>move to location near other birds in the sky (left hand side of screen)</td>
</tr>
<tr>
<td>1</td>
<td>***</td>
<td>雨伞 – 旋转, 移动</td>
<td>2</td>
<td>move to location above the table (in the correct layer)</td>
</tr>
<tr>
<td>3</td>
<td></td>
<td>桌子 – 旋转, 移动, 层移动</td>
<td>1</td>
<td>move to location below the umbrella (left side of the screen)</td>
</tr>
<tr>
<td>***</td>
<td>***</td>
<td>野餐篮 – 移动</td>
<td>1</td>
<td>move to location near right arm</td>
</tr>
<tr>
<td>***</td>
<td>***</td>
<td>手臂 - 移动</td>
<td>1</td>
<td>move to a location near body</td>
</tr>
<tr>
<td>***</td>
<td>***</td>
<td>躯干 - 移动</td>
<td>1</td>
<td>move to a location right of the table</td>
</tr>
<tr>
<td>***</td>
<td>***</td>
<td>躯干 - 移动</td>
<td>1</td>
<td>move to a location right of body</td>
</tr>
<tr>
<td>***</td>
<td>***</td>
<td>头 – 移动</td>
<td>1</td>
<td>move to a location near body</td>
</tr>
<tr>
<td>***</td>
<td>***</td>
<td>头2 – (正确位置) 移动</td>
<td>1</td>
<td>move to a location near torso</td>
</tr>
<tr>
<td>3</td>
<td></td>
<td>帽子 – 旋转, 移动, 层移动</td>
<td>1</td>
<td>rotate hat, move to a location near head2, move to correct layer</td>
</tr>
<tr>
<td>***</td>
<td>***</td>
<td>左臂 - 移动</td>
<td>1</td>
<td>move to location near torso</td>
</tr>
<tr>
<td>***</td>
<td>***</td>
<td>右腿 - 移动</td>
<td>1</td>
<td>move to location near body</td>
</tr>
<tr>
<td>***</td>
<td>***</td>
<td>左腿 - 移动</td>
<td>1</td>
<td>move to location near body</td>
</tr>
<tr>
<td>***</td>
<td>***</td>
<td>左腿 - (正确位置) 移动</td>
<td>1</td>
<td>move to near torso</td>
</tr>
<tr>
<td>***</td>
<td>***</td>
<td>弯曲右腿 – 移动</td>
<td>1</td>
<td>move to correct location near torso</td>
</tr>
<tr>
<td>***</td>
<td>***</td>
<td>腿 - 移动</td>
<td>1</td>
<td>move to correct location near body</td>
</tr>
<tr>
<td>2</td>
<td></td>
<td>上衣 – 旋转, 移动</td>
<td>1</td>
<td>rotate right side up, move to correct location</td>
</tr>
<tr>
<td>2</td>
<td></td>
<td>衬衫 – 移动, 层移动</td>
<td>1</td>
<td>rotate right side up, move to correct location</td>
</tr>
<tr>
<td>2</td>
<td></td>
<td>绿色短裤 – 移动, 层移动</td>
<td>1</td>
<td>move to correct location, above legs</td>
</tr>
<tr>
<td>2</td>
<td></td>
<td>粉色短裤 – 移动, 层移动</td>
<td>1</td>
<td>move to correct location, above legs</td>
</tr>
</tbody>
</table>

111
David Lewis is a doctoral candidate at the University of South Florida, Tampa, FL. He also is serving as an Instructional designer at Florida International University, in Miami, where he supports faculty and staff in the design and development of web-assisted, hybrid and fully-online courses. David has supported faculty and professionally designed and developed web-based course materials since 1995.