A Comparison of Audio-Visual Animated Lessons to Equivalent Static Graphics

Adam Bronsther

A thesis presented in partial fulfillment of the requirements for the degree of Master of Arts

Department of Psychology

Carleton University

Ottawa, Ontario

January, 2004

Your file Votre référence
ISBN: 0-612-93891-3
Our file Notre référence
ISBN: 0-612-93891-3

TABLE OF CONTENTS

Acknowledgements

Dick Dillon, you truly are the measure by which all supervisors should be measured. You accepted me as your student without a second thought and ever since you have shown me nothing but commitment, dedication and guidance that far exceeds any expectations any student could possibly have of a supervisor. Thank-you.

James Zdralek, thank-you for helping me program my experimental materials. Without your assistance I might still be toiling away in front of my computer trying to get it to work.

Cathy Dudek your initial optimism and enthusiasm inspired me to believe that several weeks worth of data could be collected in a week. Thank-you for providing that initial spark that sent us both on our merry way to Masters completion.

The HOT Lab gang, thank-you for providing a constant stream of support, suggestions and encouragement throughout the whole process.

Finally, to my family, thank-you for enduring the grouchy moods and procrastination; your support has meant a lot to me as I worked my way through this degree.

Abstract

This study replicated tutorial materials created by Mayer et al., (Mayer, Heiser, & Lonn, 2001b) to compare animated graphics to static graphics containing narrations, text, and narrations plus text. Participants were 90 primarily undergraduate students. Participants viewed a lesson on the formation of lightning and were asked to answer questions that measured factual learning and transfer. Animations with narrations plus text resulted in better factual learning performance than animations with text only. Animations with narrations enhanced transfer performance, but not factual learning, when compared to static graphics with narrations, but this finding must be interpreted with caution. Participants assigned to text only conditions spent significantly longer waiting at the end of lesson segments than participants in narrations only conditions. Possible interpretations of the results, methodological issues and future research were discussed.

A Comparison of Audio-Visual Animated Lessons to Equivalent Static Graphics

Research on the effectiveness of animation in learning has generally been unconvincing. The majority of research cited and conducted by Wong (1993) and Hutcheson (1997) indicated no significant advantage of animation over the use of comparable still images on participants' ability to learn (retain) facts. The studies that did demonstrate advantages of animations over static graphics were criticized for questionable methods and statistical interpretation.

After reading these studies one might be pessimistic about the merits of animation and multimedia for learning and training applications. However, the notion that multimedia allows for better learning persists. Prior to and since Wong and Hutcheson's research, a series of studies conducted by Mayer and his colleagues (Mayer & Anderson, 1991; Mayer, Bove, Bryman, Mars, & Tapangco, 1996; Mayer & Moreno, 1998; Moreno & Mayer, 1999; Moreno & Mayer, 2000) have, to varying degrees, demonstrated support for the effectiveness of animation in learning. Mayer converted the results of these studies and others into what he referred to as the "Seven Principles of Multimedia Learning" which are design guidelines for the formation of effective multimedia learning tools. However encouraging Mayer's results are, one serious criticism can be levied against them: Mayer's studies have never sought to compare the effectiveness of animations against corresponding static graphic counterparts.

The goal of the current research is to replicate and extend Mayer's research by comparing an animated lesson on the formation of lightning to an equivalent static graphic version of the same material.

Review of Literature

Before exploring the literature, some definitions are provided. Multimedia refers to the combination of multiple media. Forms of media can include, but are not limited to sound and music, still graphics, animated graphics, written text, narrated (spoken) text, video, and even hyperlinks. Multiple media may also involve other senses such as smell or touch through the output of olfactory stimuli such as the smell of seawater during a virtual tour of the ocean or through sensations like resistance or vibration. In this research, multimedia is limited to graphics, voice narrations and written text. The graphics are either static, meaning they are stationary; or they are animated, meaning parts of the images appear to move across the screen as though in motion or to change image without moving. Also the terms "comparable" and "equivalent" are used to describe animations and static graphics that are designed to be equivalent in appearance and content. The only difference between comparable graphics in this thesis is that one is animated while the other remains static. For example, while an animated lesson contains moving pictures and a static graphics lesson has still pictures; both lessons contain verbatim factual content whether presented as text or narrations. Learning measures are discussed in terms of factual recall, recognition, and transfer. Factual recall occurs when a learner is able to remember facts presented in lesson materials by generating the facts. Factual recognition occurs when a learner is able to recognize a fact that has been presented in lesson materials, such as when responding to multiple-choice questions. Transfer of learning occurs when the learner is able to answer a unique question based on material they were presented, but that was not directly taught as part of the lesson. Thus, in this study, transfer refers to the application of learned material to new situations, while factual learning is the recognition of facts presented.

Studies by several researchers are discussed including Rieber (1990; 1991), Wong (1993), Hutcheson (1997) as well as Mayer, Moreno (Mayer et al., 1998; Mayer et al., 2001b; Moreno et al., 1999) and Tversky, Betrancourt and Morrison (2002).

Rieber (1990) studied the relation between type of lesson (static graphics, animated graphics, and no graphics) and type of practice (behavioural, cognitive, and no practice) and their effect on learning with elementary school children. The learning materials dealt with aspects of Newton's laws of motion. Each lesson was divided into four segments, with each segment dealing with a different aspect of Newton's laws. For example, the first part dealt with who Newton was and his discovery of Newton's first law; the second part demonstrated that once an object is set in motion, an equal force in the opposite direction is required to stop it. The relevant dependent measure in this study was performance on a 35-question, multiple-choice test.

A no graphics condition contained text only. The static graphics condition included the textual lesson used in the non-graphics condition coupled with static graphics placed throughout the lesson to demonstrate the concepts and rules. The application of forces on objects was demonstrated by an image of a foot kicking a ball and motion and trajectory were indicated through the use of arrows and path lines.

The animated graphics condition was similar to the static graphics condition in that it used a foot kicking a ball to demonstrate the application of forces, but in this condition the foot and ball were animated such that the ball was kicked twice to the right across the screen by a foot and then stopped by two kicks to the left. This condition also included the textual material.

Rieber also included three practice conditions that he called behavioural practice, cognitive practice and no practice. Rieber defined behavioural practice as five multiple-choice questions on the lesson segment that had been viewed, presented after each of the four lesson segments. Rieber did not indicate the relation between the behavioural practice questions and the post test questions. After the behavioural practice questions, participants were provided with feedback about the accuracy of their answers.

Cognitive practice involved a computerised simulation of a space ship. By simulating the effect forces would have on the motion of the space ship, the simulation was designed to provide participants with a practical application of Newton's laws that had been demonstrated by the foot kicking the ball. Participants were given increasing control of the ship after each lesson segment. For example, after the third lesson segment, participants were allowed to increase or decrease the speed of the ship and after the fourth lesson segment they were allowed to apply forces to the ship in 90 degree increments. A no practice condition involved participants progressing through the lesson uninterrupted.

Rieber observed a graphic type by practice interaction on the multiple-choice test such that animations when coupled with behavioural practice were more effective than static graphics when coupled with behavioural practice. For cognitive practice and no practice, animations were not superior to static graphics. Participants who received either form of practice took significantly less time to answer post test questions than those with no practice but there were no differences between static and animated graphics on time to answer questions.

In a follow up experiment with elementary school children, Rieber (1991) studied the effects of lesson type (static, or animated) on intentional and incidental learning and preference of material. The lessons and practice were the same as those used in Rieber, 1990. The relevant dependent measure in this study was performance on a 24-point multiple-choice test that measured both intentional and incidental learning.

According to Rieber, incidental learning is "learning those objectives that are not directly taught but only implied through contextual cues provided in several of the animated visual displays" (Rieber, 1991). Accordingly, if participants demonstrated that they understood, through performance on the multiple-choice test, a concept of Newton's second law that was not directly explained, then, according to Rieber, incidental learning occurred.

Incidental learning occurred significantly more as a result of an animated lesson versus a static version of the same lesson. Also, intentional learning occurred significantly more as a result of animated lessons versus static lessons, but this effect was not as strong as for incidental learning. Students in the animated lesson also performed significantly worse than students in the static lesson on two questions pertaining to Newton's law of gravitation. Rieber suggested that errors of interpretation are possible with animations when students are not provided with direct support (such as a teacher) while learning.

Results from the two Rieber studies suggest that the use of animation has some benefits in learning that static graphics do not have, but that these effects may be mediated by other factors including the type of practice and the type of material tested.

Several aspects of Rieber's research are problematic. One problem was identified by Tversky and her colleagues (Tversky, Morrison, & Betrancourt, 2002). Specifically, they determined that Rieber's lessons did not contain comparable content between the animations and static graphics. Rieber's animated lesson about Newton's law of equal and opposite forces showed a foot kicking a ball in one direction. The lesson also showed a foot kicking the same ball in the opposite direction to stop the ball's movement. This animation aided in explaining the concept of inertia. The static version of this lesson showed the movement of the ball, but failed to demonstrate that a kick both started and stopped the ball's movement. Information about this fact was available only from the accompanying text in the static condition rather than as part of the images, suggesting that learners had to work harder to receive the same information as those in the animated condition (Tversky et al., 2002). Furthermore, Rieber's reliance on fourth and fifth grade students limits conclusions about how these findings generalize to a broader population. It is conceivable that the effects of animations are more or less pronounced for learners of different ages.

Wong (1993) compared the effects of animation and comparable static graphics on learning the statistical concept of Analysis of Covariance (ANCOVA). This topic is considered abstract and quite difficult by students in statistics courses. Wong wanted to measure the effect of her tutorials on learning ANCOVA concepts as well as to determine if animations created using a set of proposed design guidelines would be more effective than those that were not. Undergraduate and graduate Psychology student participants were assigned to one of three conditions: a guideline consistent animated tutorial, a guideline inconsistent animated tutorial, and a static graphic tutorial. Wong's guidelines were formulated based on her experience creating and revising the materials for her experiment, through observation of other tutorials, as well as through support from the literature on multimedia and animation. The guidelines were meant to address common problems that were experienced during the creation of the tutorials. Examples of some of the guidelines include "avoid overuse of blinking", "avoid overuse of transition effects", and "animate essential objects only." Thus, the guideline consistent tutorial applied the guidelines generated by Wong while the guideline inconsistent tutorial was created with deliberate violations of the guidelines.

All three conditions contained the same lesson material and used static graphics and text to teach the concept. The animated lessons also contained animated graphics. In the static graphics condition, a series of individual graphics were used with motion and movements indicated by arrows. The static and animated graphics used the same start and end graphics, but the animated graphics replaced the arrows and lines with animations. The dependent measures of interest were performance on a quiz consisting of factual recognition, and transfer questions. Questions involved multiple-choice, circle the correct response, matching, and drawing. At the end of the session, participants were also asked to offer subjective ratings of the lessons pertaining to satisfaction and other issues.

The guideline consistent animation group did not perform any better than the static group on either learning or transfer questions. There were also no differences between the two groups on time to complete the tutorial or the quiz. However, the guideline inconsistent animation group performed poorer on factual questions and took longer to complete tutorials than the static graphic or guideline consistent groups. Wong found no differences between the groups on ratings of satisfaction of the lesson types. These results suggest that well-designed, comparable static graphics are as effective as good animations for learning, but animations that do not help to explain the concepts are actually detrimental.

Wong's text appeared distally from the animations, which is a violation of Mayer's *Spatial Contiguity Principle* (Mayer & Moreno, 2002) that suggests that text should be placed proximally to the animation that the text describes. Since Wong's text appeared above the graphs in all conditions, it is possible that participants' attention was split between the animation and the text such that performance may have been affected. Also, related to the placement of the text, is the fashion in which it was presented. Text appeared in a box, which separated the text from the rest of the lesson. Participants may have been visually drawn to the text since it took up a large amount of screen space and was enclosed in a box.

As an extension to Wong's study, Hutcheson (1997) added voice narrations that served to cue important information in Wong's animations. The narrations did not replace the text but, rather, highlighted important aspects in the text and animation as it happened. Hutcheson also modified the quiz. Both Wong's and Hutcheson's quizzes contained questions that presented visual materials in the form of graphs and distributions. However, some of Wong's questions addressed material covered in textual material that was not covered graphically, therefore some questions did not probe concepts that the animations in the lessons demonstrated. Also, some of Wong's questions presented response alternatives in text form when the material had been presented in graphic form, possibly influencing recognition performance. Hutcheson asked questions only on material that had been presented in graphical form and created questions with response alternatives that were images rather than text.

Conditions in this study were an animated lesson and a static graphics lesson about Analysis of Covariance. Both lessons made use of static graphics, narrations and text. Graphics, either animations or comparable static images, were used to convey relations between variables in a data set and how the statistical adjustment affected the results obtained. Following Wong, text was placed in a box at the top of the screen. The text and narrations described the particular concept being demonstrated. Dependent measures included accuracy on the quiz, time to complete the tutorial, time to complete the quiz, time spent on each page of the lessons, and number of pages accessed by the participants.

There were no significant differences between participants in the animation-tutorial group and static-tutorial group on accuracy of results. Participants spent significantly more time completing the animated tutorial and spent more time on animated tutorial pages than those in the static tutorial group. These results suggest that well-designed animations that illustrate important learning concepts are no better than a series of comparable, well-designed static graphics but that people spend more time learning with the animations.

Some of the problems with Wong's study are present in Hutcheson's. For example, presenting text distally from the animations violates Mayer's Spatial Contiguity Principle. Hutcheson used voice narrations that coincided with important details of the animated tutorials. Having animations, narrations and on-screen text is a violation of Mayer's *Redundancy Principle* (Mayer et al., 2002), which states that students learn more deeply from animations and narrations than from animations, narrations, and on-screen text.

Research by Mayer and Moreno that appeared after the Wong and Hutcheson studies provides new possibilities as to why Wong and Hutcheson found no advantages of animations over corresponding still graphics. Three multi-part studies carried out by Mayer and Moreno (Mayer et al., 1998; Moreno et al., 1999; Mayer et al., 2001b) provide the background and inspiration for the current research. Each study used an animated tutorial on the formation of lightning to evaluate under what circumstances retention and transfer occur, but did not compare animations with static graphics. The experiments were used to further a set of design guidelines for the use of narrations and animations in multimedia tutorials.

The first two experiments (Mayer et al., 1998) sought to determine how animations with synchronized narrations versus synchronized on-screen text would affect retention and transfer. One experiment used an animated tutorial on the formation of lightning while a second tutorial depicted the operation of an automobile braking system. In both experiments, participants were college students. Participants in both experiments were assigned to either an animated tutorial featuring synchronized narrations, which offered voice descriptions coinciding temporally with the topics being animated, or to an animated tutorial featuring on-screen text that coincided temporally with the topics being animated. From the Mayer et al. (1998) description, it seems that the text matched the narrations verbatim and when the animations stopped, the text disappeared.

In both experiments, results relevant to my study were that participants in the synchronized narrations condition recalled significantly more facts and gave more correct transfer answers than those in the synchronized on-screen text condition.

Moreno et al., (1999) performed an additional experiment that made use of the lightning animation tutorial. The goal of this experiment was to determine how the placement of on-screen text and timing of on-screen text and narrations would affect retention, transfer and matching test results. Participants were college students, and dependent measures were the same as those used in the previous study.

Two of the three conditions compared animations with concurrent narrations to animations with concurrent text that was located close to the animation. Participants assigned to the narrated condition scored significantly higher on recall and transfer tests than did participants in the text condition.

Mayer et al., (2001b) followed their earlier work with two experiments that again made use of the lightning tutorial. Both experiments used college students as participants and among other things, measured learning and transfer. In the first experiment, two of the four conditions included one group that received animations with concurrent narrations. A second group received narrations plus identical, concurrent on-screen text. Students in the narrations plus on-screen text condition retained significantly fewer facts and produced fewer transfer solutions than those in the narrations condition.

Mayer hypothesized that the superiority of narrations to text and narrations plus text results from the first experiment were indicative of a *redundancy effect* in which the addition of simultaneous on-screen text to a narrated animation could lead to poor performance on recall and transfer. A second experiment with three conditions addressed this issue. One of the relevant conditions was a no-text group (animations, narrations, no text), and the second relevant condition was a group with on-screen text that matched the narrations verbatim (animations, narrations, verbatim text). Participants in the narrations with no-text condition performed significantly better on measures of recall and transfer than did participants in the narrations plus text conditions.

Mayer suggested a *split-attention hypothesis* that says "when words are presented visually, learners must split their visual attention between the on-screen text and the animation, thereby failing to adequately attend to some of the presented material" (Mayer et al., 2001b). It is not surprising, that narrations describing what was happening while the animation occurred resulted in better performance than text presented while the animation occurred. People were able to listen and view the animation at the same time, but people were not able to view the animation and read the text; or view the animation, listen to the narrations and read the text at the same time.

This series of studies seems to clearly indicate that narrations during animations are superior to text and superior to narrations plus text during animations. However, there are problems with the studies. In addition to the fact that no comparisons were made of animations to comparable static versions of the tutorials, another problematic aspect of Mayer's series of studies deals with the transfer questions where the detrimental effect of text was substantial. The scoring of responses to transfer questions seems arbitrary and inconsistent. Mayer presented participants with four transfer questions that included "What could you do to decrease the intensity of lightning", and "Suppose you see clouds in the sky, but no lightning. Why not?" He also offered suggestions for acceptable answers to these questions such as "removing positive ions from the ground" for the first question and "the tops of the clouds might not be high enough to freeze" for the second question. Moreno et al., (1999) also described *unacceptable* answers to these questions. An example of an unacceptable answer to the first question included "removing trees and tall objects from the ground" and "the cloud was not a rain cloud" for the second question. It is unclear why some responses are acceptable and others not. For example, removing tall objects from the ground seems a more plausible solution to preventing lightning than removing positive ions from the ground. The concern is that learners did not have any guidance for what constitutes an acceptable or unacceptable answer and the scoring seems arbitrary so performance on transfer problems may not be reliable or valid.

These and other studies also demonstrated circumstances for the effective use of animated graphics, narrations and text. Narrations should be presented concurrently with animations instead of preceding or following animations (Mayer et al., 1991; Mayer & Sims, 1994; Mayer, Moreno, Boire, & Vagge, 1999). If text is used, it should be placed close to the animation (Moreno et al., 1999). Animations with narrations should be presented without the addition of extra details such as music, sounds, and interesting but irrelevant information (Mayer et al., 2001b; Moreno et al., 2000). Some evidence suggests that learners with low spatial ability derive less benefit from animations and narrations than those with high spatial ability (Mayer et al., 1994).

Mayer and associates have also demonstrated that static graphics are better than no graphics[1] (Mayer et al., 1996), and how to effectively use static graphics. Static graphics should contain text captions highlighting key points in the lesson material and brief textual summaries

[1] Rieber (1990) did not find that static graphics are better than no graphics.

instead of full text passages (Mayer et al., 1996). Presenting static graphics alone with no accompanying captions results in poor learning performance (Mayer et al., 1996).

Overall, Mayer's research demonstrates that well-designed still illustrations and well-designed animations both help learning. There are ideal conditions under which well-designed illustrations and animations will be effective in helping learners understand scientific explanations.

Tversky et al. (2002) conducted a critical literature review of some of the relevant literature regarding the use of static images, animations and interactive multimedia in educational lessons. The main motivation behind the review was the concern that research on the effectiveness of animation in learning must be based on direct comparisons of animations and equivalent static image counterparts. Furthermore, not only should the experimental materials be equivalent, but the experimental procedures must also be comparable. For example, in a study comparing animations to static graphics, if one group of participants received interactive animated graphics then participants receiving static graphics should also receive interactive static graphics. Tversky et al. offered examples in the literature where researchers violated these principles in their experiments. In experiments where the conclusion was that animated tutorials were more effective than static tutorials, the researchers often offered participants better designed animations, or animations that included information that was not present in the static lessons, as in Rieber's study discussed earlier. Tversky et al. highlight studies from other researchers where non-equivalent comparisons were made between animations and static graphics including studies from Baek and Layne (1988), Large (Large, Beheshti, Breuleux, & Renaud, 1996), and Park and Gittelman (1992).

Baek and Layne (1988) presented learners with materials about differences in speed for rate, time and distance problems. Animated materials consisted of pairs of dashed lines that moved in proportion to speed. Static materials consisted of a table that listed two sets of distances, times and speeds. The researchers claimed benefits for animations over static graphics on learning measures. However, most people would not consider tables to be graphics and the material in tables was certainly not comparable to the material in the animations. A similar criticism was levied against Large et al., (1996) because in their study comparing animations to static graphics using lesson materials about the circulatory system, the animated materials included blood pathways but the static graphic materials did not. In both the Baek and Layne and Large et al. studies, learners were provided with cues or information about the learning materials in the animated conditions that were not obvious or present in the non-equivalent static condition. Another example described by Tversky et al. comes from research by Park and Gittelman (1992). The researchers indicated better performance on learning scores from animated lessons on the operation and troubleshooting of an electronic circuit, yet the animated materials showed detail that the static graphic materials did not. Specifically, the animated materials showed the fine detailed behaviours, such as changes in state, of the circuit while the static graphic materials only demonstrated the spatial relationships between components.

Tversky et al. also described studies where researchers employed non-equivalent procedures, such as (Schnotz, Bockheler, & Grzondziel, 1999) and (Hegarty, Quilici, Narayanan, Holmquist, & Moreno, 1999) who included elements of interactivity or prediction in their animated conditions but not in the static graphics conditions. Interactivity can be defined as components in lesson materials that allow the learner to take control over the presentation of the materials. Examples of interactive elements can be as simple as including "proceed" and "repeat" buttons or more complicated such as hyperlinks that learners click to access more information or direct manipulation of objects such as rotating and breaking apart a motor to see the internal structure. Interactivity promotes "learning by doing" since the learner is more likely to remember information if they have taken an active part rather than passively viewing materials (Tversky et al., 2002). Prediction occurs when the lesson materials encourage learners to make guesses about outcomes and view results in the form of graphics or text to see if their guesses were correct. For example, learners could be asked to make predictions about the behaviour of computer algorithms or weather patterns. Tversky et al. state that both interactivity and prediction are known facilitators of learning, therefore, the inclusion of them in animated conditions but not in static graphics conditions is methodologically poor and makes it difficult or impossible to make definitive statements about animations.

Goals for Research

There were several goals for this research. One was to properly compare equivalent static and animated tutorials on their effectiveness for facilitating factual learning and transfer. Mayer's research on animations demonstrated that, under proper circumstances, animations can be effective in facilitating both factual recall and transfer performance. Mayer's research also demonstrated that static illustrations can be effective in facilitating recall and transfer performance. However, Mayer never compared his static illustrations to equivalent animated conditions to see if differences existed in effectiveness.

Hutcheson (1997) and Wong (1993) failed to find evidence for the use of animations versus static images. Wong compared animations with text to static graphics with text and found no benefits for animations on performance. Hutcheson compared animations with narrations plus text to static graphics with text and still found no benefits for animations on performance.

Mayer and Moreno (Mayer et al., 1998; Mayer et al., 2001b; Moreno et al., 1999) have consistently found that animations with narrations but no text result in better performance than narrated animations that include text. This study attempted to reconcile these conflicting results by comparing narrated animations and comparable narrated static graphics without the presence of text.

Based on this background information, several hypotheses were generated. Hypothesis 1 was that animations with voice narrations but no text would result in better factual learning and transfer than static graphics with voice narrations but no text. Hypothesis 2 was that animations would not result in better factual learning and transfer than static graphics if both animations and static graphics were accompanied by text only or by narrations plus text. Hypothesis 3 was that, for animations, narrations would result in better factual learning and transfer than text only. Hypothesis 4 was that for static graphics, narrations would result in better factual learning and transfer than text only. Hypothesis 5 was that for animations, text only would result in better factual learning and transfer than narrations plus text. Hypothesis 6 was that for static graphics, text only would result in better factual learning and transfer than narrations plus text. These comparisons have never been made in a properly controlled study.

Method

Experimental Design

The experiment was a 3x2 completely between-groups design with participants randomly assigned to one of the six conditions. Static and animated conditions were crossed with text only, narrations only, and text plus narrations. Dependent measures included performance on recognition and transfer questions as well as how much time participants spent waiting at the end of an instructional segment before clicking the "proceed" button.

Participants

Ninety participants[2] were recruited from the pool of undergraduate psychology students registered in the summer term and from the population of students on the Carleton University campus based on availability meaning that any willing participant was allowed to participate. In addition to Introductory Psychology signup sheets, notices were posted across the campus and an e-mail was sent to a distribution list of Carleton University students who indicated they were interested in knowing about paid activities occurring on campus.

Apparatus

All sessions were run on AMD® Athlon based computers running at 1.8 GHz or Intel Pentium 4® class computers at 2.4 GHz, running Microsoft Windows 2000 Service Pack 2 on 17" monitors with resolution set to 1024x768 at 32 bit colour. Participants interacted with the computers via a mouse.

Materials

Tutorials

Using Moreno and Mayer's research (Moreno et al., 1999) as guidance, tutorials that explained how lightning forms were generated. Tutorials took the form of either animated or static graphics lesson with three variations of each. Tutorials contained narrated descriptions, textual descriptions, or textual and narrated descriptions. All tutorials were approximately three

---

[2] Results for two participants were replaced during the experiment because their performance on the multiple-choice questions. Both participants were assigned to animations with on-screen text. One participant scored 2 out of 19 (10%) while the other participant scored 3 out of 19 (16%). Both of these scores are well below chance (25% with four alternatives). The next lowest score was 8 out of 19 (42%), which occurred only once, suggesting that the replaced participants did not take the lesson seriously.

minutes in length. All tutorials were broken down into 16 segments, with each segment addressing a key concept in the formation of lightning. These segments are listed in Appendix A. All tutorials allowed participants to advance to the next segment at their own pace by pressing the "proceed" button in the lower right hand corner of the tutorial screen. Participants could not go back and view a segment after it completed.

The animated tutorials used moving graphics to demonstrate air flow, movement of charged particles and steps of lightning bolt leaders. The static tutorial used arrows to demonstrate movement. All tutorials were created using Macromedia Director MX (Macromedia, 2002). Figure 1 demonstrates two images from the static tutorial.

Figure 1. Two images from the static lesson on the formation of lightning.

In animated tutorials, screens, with movement instead of arrows, were identical to the content in the static tutorial.

The narrated tutorial presented participants with 16 distinct audio statements. For example, statement one said "Cool moist air moves over a warmer surface and becomes heated", and statement 11 said "A stepped leader of negative charges moves downward in a series of steps. It nears the ground." The spoken statements coincided temporally with the animations that occurred on-screen.

The on-screen text tutorial presented participants with 16 distinct textual statements. These statements matched the narrated statements verbatim and were placed proximally to the concept being taught as illustrated in Figure 2.

Figure 2. Image demonstrating on-screen text.

The text plus narrations tutorial combined both the narrations from the narrations tutorial and the identical on-screen text from the text tutorial. Text remained visible on-screen until participants clicked the "proceed" button. In the text and narrations condition, this meant that the text remained visible after animations and voice narrations ended. In the narrations condition, the graphic was available for viewing after the narrations stopped, but there was no text or voice.

Questions

Participants received a pre-session questionnaire, and a 19 item multiple-choice recognition test that consisted of 16 factual learning questions and three transfer questions. Questions were part of the tutorial program and were completely automated so that participants clicked items on-screen to answer the questions. Performance on the questions was graded by the program with all results output to a text file. Participants were not shown their scores unless they asked to see the results.

The pre-session questionnaire, which is presented in Appendix B, asked participants about their meteorology knowledge, and a self-assessment rating of meteorology knowledge.

The recognition test, presented in Appendix C, consisted of 19 recognition (multiple-choice) questions, with questions one through 16 being factual learning questions and 17 through 19 transfer questions. Transfer of learning occurs when the learner is able to answer a unique question based on material they were presented, but that was not directly taught as part of the lesson. In this study, an example of a transfer question is "Which of the following could conceivably decrease the intensity of lightning?" The participant was presented with four options including "Increase positive charges in the air", "Remove negative charges from the ground", "Increase the temperature of the air near the Earth's surface", and "Remove positive charges from the ground."

Procedure

Participants filled out the informed consent form presented in Appendix D, and were then instructed about the procedure of the experiment. All participants were told that the study was completely automated and that they would be participating by themselves. They were told there would be a series of questions at the end of the tutorial, that they should not leave any questions blank, and that they should take the lesson seriously as they would any other type of lesson. Following this, participants were assigned to an experiment room where they viewed one of the tutorials, depending on condition they were assigned to. Participants who heard narrations were instructed on-screen to put on headphones. On screen instructions told them to push the "Proceed" button to begin the lesson.

After viewing the tutorial, on-screen written instructions told participants they would be presented with 19 multiple-choice questions. Participants were told they could take as much time as they needed to answer the questions, to use the "proceed" button to advance to subsequent questions, and that they must answer all questions. Sessions took about 20 minutes.

Results

Results were recorded for prior knowledge and for the three dependent measures that consisted of performance on multiple-choice questions, and time participants spent waiting at the end of each segment before clicking a button to proceed to the next segment. Prior knowledge was assessed by asking participants to place check marks next to seven statements such as "I know what a cold front is", and "I know what a low pressure system is." A self-assessment rating was obtained by asking participants to rate on a five point scale their meteorology knowledge, with one being "very little" and five being "very much."

Performance on the 19 post-lesson recognition questions was divided into a factual learning component and a transfer component[3]. The factual learning component consisted of the first 16 recognition questions that tested knowledge of facts presented in the tutorial. With a

---

[3] Analyses were also performed on the total score out of 19 that combined the learning and transfer scales. Results for the total score, presented in Appendix E, were consistent with results for the separate scores and provided no information that was not already available from the separate scores.

correct or incorrect score on each question, possible values ranged from 0 to 16. These were converted to percent correct. Three transfer questions required participants to apply understanding of the material that was presented in the tutorial to situations that were not covered in the tutorial, but that would be possible to answer if the participant understood the material. Scores on the transfer measure, which ranged from 0 to 3, were also converted to percent correct.

The additional-time dependent variable was the difference in time between the end of a lesson segment and the time at which participants clicked the "proceed" button to advance to the next segment. It is assumed that the period immediately after a segment was presented might reflect time thinking about the material just presented and might be influenced by the tutorial conditions. Additional time was measured by the tutorial software in milliseconds and converted to seconds then summed for all 19 segments to get a total time measure.

Reliability of factual learning and transfer

A reliability analysis using Cronbach's alpha was performed to assess whether the 16 factual learning questions and three transfer questions were measuring unidimensional latent constructs. The analyses looked at consistency among the factual learning and transfer questions, respectively. A high degree of inter-correlation would yield a large alpha value which would indicate that the individual questions were reliably measuring factual learning and transfer. Cronbach's alphas were .56 for factual learning and -.25 for transfer, considerably below the .70 that is usually taken as an indication of adequate consistency. The low alpha value for factual learning and negative value for transfer indicate that the measures for factual learning and transfer are complex and multidimensional .

Data Analyses

Separate 2 by 3 completely between-groups factorial univariate ANOVAs were performed on the two measures of prior knowledge and the three dependent variables.[4] These analyses had two levels of graphic type of tutorial (static or animated) as one factor and three levels of description modality for descriptions of what was presented in each segment (text description only, narrations only, or text plus narrations) as levels.

Prior Meteorology Knowledge

Participants rated themselves on prior knowledge of meteorology. This information was used to ensure that groups were equivalent on knowledge of meteorology. The ANOVAs, means and standard deviations, and graphs for meteorology knowledge are presented in Appendices F through H. Main effects of description modality and graphic type, and the interaction between them were not significant for prior knowledge and self rating measures of meteorology, all Fs < 1. These results indicate that the groups were not knowledgeable about meteorology and the groups were equivalent on knowledge about meteorology before the tutorials were presented.

Overview of factual learning and transfer performance

Separate ANOVAs on factual learning and transfer indicated that there were no main effects of graphic tutorial type, largest $F(1,84) = 3.18$, $p > .07$, $\eta^2 = .036$; no main effects of description modality, largest $F(2,84) = 3.01$, $p > .05$, $\eta^2 = .067$; and no interaction between graphic type and modality, largest $F(2,84) = 1.39$, $p > .25$, $\eta^2 = .032$. The ANOVAs, means and standard deviations for factual learning and transfer are presented in Appendices J through M.

---

[4] A multivariate analysis of variance (MANOVA) was performed using all three dependent variables, but provided no information that was not already known from the separate univariate analyses. The MANOVA Summary Table is presented in Appendix I.

These statistical tests suggest an interpretation of the results that specifies that none of the treatments had any effect, alone or in combination, on factual learning or transfer performance.

Despite these non-significant effects, I have addressed the six hypotheses of the study as planned, non-orthogonal comparisons. All of these comparisons are found within the non-significant interactions between graphic type and description modality. Planned non-orthogonal comparisons can be made even if the interaction is not significant but they require adjustments to the alpha level for comparisons among the description modalities, so alpha is .05/3 for each of these comparisons. Because there are only two levels of the graphic type independent variable, alpha is .05 for comparisons between animated and static graphics. Planned non-orthogonal comparisons are used to provide the strongest possible chance to find statistically significant results for tests of predictions from Moreno (1999) and Mayer's (2001b) research, which are generally not supported.

Factual learning Performance

Figure 3 demonstrates the interaction within which the planned comparisons for factual learning are found. In figure 3, the vertical axis represents percent correct on factual learning questions, while the horizontal axis represents description modality. Graphic type is shown by two lines for animations (solid line) and static graphics (dashed line).

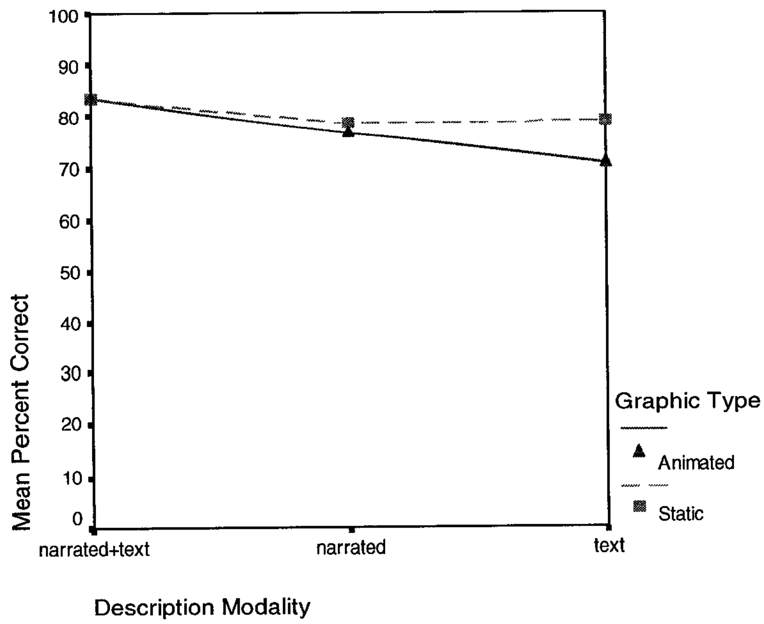

Figure 3. Interaction between description modality and graphic type for factual learning questions

Description Modality for Animations and Static Graphics

Animations

Factual learning performance for animations is shown as the solid line in figure 3. Based on research by Moreno (1999) and Mayer et al.(2001b), it was expected that animations with narrations would result in the best factual learning, followed by animations with text while animations with narrations plus text would hinder performance. The simple main effect of description modality for animations was significant, $F(2,84) = 3.14$, $p<.05$, $\eta^2 = .07$ indicating that there were differences in factual learning depending on description modality when animated tutorials were used. The surprising finding for the samples used in this experiment was that animations with narrations plus text resulted in the best factual learning, not the worst. With animated tutorials, factual learning with narrations plus text was significantly better than factual learning for animations with text only, $p<.02$. Animations with narrations plus text did not differ from animations with narrations, $p > .17$, and animations with narrations did not differ from animations with text only, $p > .26$.

Static Graphics

The simple main effect of description modality when static graphics were used was not significant, $F < 1$. Thus, there is no evidence that the dotted line representing factual learning performance with static graphics tutorials in Figure 3 is not flat. Description modality did not make any difference when static graphics were used.

Animations with Narrations vs. Static Graphics with Narrations

The main hypothesis of this research was that animations with narrations would be more effective than static graphics with narrations for factual learning measures. Wong (1993) and Hutcheson (1997) failed to find differences between animations and static graphics when coupled with text or text plus narrations but Moreno (1999) and Mayer et al., (2001b) have demonstrated that the ideal condition for learning from animations is when they are coupled with narrations but no text.

As shown in the middle of Figure 3, contrary to predictions, for the sample data in this experiment mean factual learning performance with static graphics was slightly higher than with animated graphics, but this difference was not significant, $p > .65$. This finding supports the work of Wong and Hutcheson who found no benefits of animations over static graphics.

Animations vs. Static Graphics: Text only and Text plus Narrations

Another hypothesis of this study was that there would be no differences between animations with text or text plus narrations versus static graphics with text or text plus narrations for factual learning. These findings can be observed on the left and right sides of figure 3. No significant differences between animated and static graphics were observed for narrated plus text, p>.99; or for text only, p>.10.

Transfer Performance

Figure 4 demonstrates the interaction within which the planned comparisons for transfer are found. In figure 4, the vertical axis represents percent correct on transfer questions, while the horizontal axis represents description modality. Graphic type is represented by two lines for animations (solid line) and static graphics (dashed line).

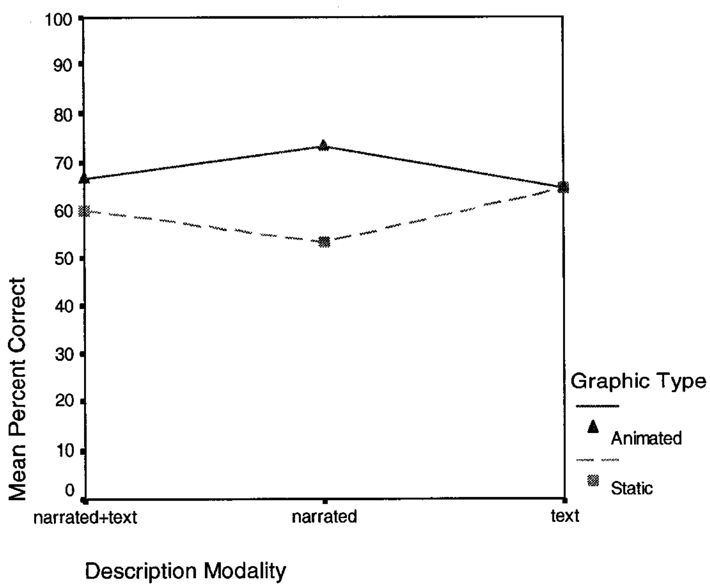

Figure 4. Interaction between description modality and graphic type for transfer questions

Description Modality for Animations and Static Graphic

Based on research by Moreno (1999) and Mayer (2001b), it was expected that animations with narrations would result in the best transfer, followed by animations with text while animations with narrations plus text would hinder performance. The simple main effects of description modality for animations and for static graphics were not significant, both Fs<1, indicating that there were no differences in transfer depending on description modality when animated tutorials or graphic tutorials were used. Description modality did not make any difference when either animations or static graphics were used.

Animations with Narrations versus Static Graphics with Narrations

As was the case with factual learning, a major issue for this study was that animations with narrations would be more effective than animations with static graphics on transfer performance. As shown in the middle of Figure 4, animations with narrations were more effective than static graphics with narrations on transfer measures, $p<.03$.

Animations vs. Static Graphics: Text only and Text plus Narrations

Differences between animations with text or text plus narrations versus static graphics with text or text plus narrations for transfer can be observed on the left and right sides of figure 4. No significant differences between animated and static graphics were observed for narrated plus text, $p>.44$; or for text only, $p>.99$.

Total Additional Time

ANOVA summary tables, means and standard deviations for the total additional time measure are presented in Appendices N through M. As was the case with the two performance measures, there was no significant main effect of graphic type, $F(1,84)=1.17$, $p>.28$, $\eta^2=.014$; and no interaction between graphic type and description modality, $F(2,84)=1.77$, $p>.17$, $\eta^2=.04$; but there was a significant main effect of description modality, $F(2,84)=8.67$, $p<.001$, $\eta^2=.171$ Means and standard deviations for the three levels of description modality are shown in Table 1.

Table 1

Means and standard deviations (in parentheses) for total additional time spent waiting

Description modality		
Text + Narrations	Narrations	Text
48.97 sec.	33.13 sec.	61.94 sec.
(29.64)	(18.64)	(31.26)

Specific predictions were not made about the amount of time participants would spend at the end of segments, so multiple comparisons using a Bonferroni correction were performed to explore the differences in total time for the significant verbal-modality effect. Thus, the alpha level of .05 was divided by three comparisons to give an alpha of .017.

Participants who received text alone spent an average of 62 total seconds waiting at the end of instructional segments before clicking the "proceed" button. This was significantly longer than participants in the narrations only condition who waited an average of 33 total seconds at the end of segments, $p<.001$. Differences between text and text plus narrations, $p> .19$, and between narrations and text plus narrations, $p> .07$, were not significant.

Summary of Results

A summary of the results based on the hypotheses for factual learning and transfer are presented in table 2.

Table 2

Summary of Results

	Result	
Hypotheses	Factual learning	Transfer
Animations + narrations better than static graphics + narrations	Not supported	Animation + narrations > static graphics + narrations
Animations + text not better than static graphics + text Animations + narrations + text not better than static graphics + narrations + text	Supported	Supported
Animations + narrations better than Animations + text	Not supported	Not supported
Static graphics + narrations better than Static graphics + text	Not supported	Not supported
Animations + text better than Animations + narrations +text	Animations + (narrations + text) > animations + text only	Not supported
Static graphics + text only will result in better factual learning and transfer than narrations + text	Not supported	Not supported

In addition, participants took longer with text only than with narrations only.

Discussion

Reliability and Effect Sizes

Given the low reliability of the factual learning and transfer measures, it is possible that what I have called factual learning and transfer are in fact complex constructs. While there are patterns in the data, what these patterns are indicative of is not clear. The following discussion should be read with this ambiguity in mind.

Effect sizes for graphic type, description modality and their interaction on factual learning and transfer were small. Whether animations or still graphics were used and whether they were combined with narration, text or narrations plus text accounted for only a small proportion of the variance in participants' factual learning and transfer performance. In terms of the impact on performance the choice of graphic type and description modality doesn't seem to have much of an impact.

Summary of Previous and Present Research Findings

Most of the background material for this research suggested that, if differences exist between animations and static graphics, they would be found in the comparison of animations with narrations to static graphics with narrations. Previous attempts failed to find differences between animations and static graphics when they were coupled with text (Wong, 1993) or text plus narrations (Hutcheson, 1997). Mayer and Moreno have consistently found that animations are most effective when coupled with narrations without text (Mayer et al., 1998; Moreno et al., 1999; Mayer et al., 2001b). The main finding for this thesis was that there are no differences between animations and static graphics for measures of factual learning and a qualified difference for transfer. By planned comparison tests, one of the three statistically significant findings in this study indicated that animations with narrations were more effective for transfer performance than static graphics with narrations, but this finding is not easily interpretable. Another significant finding was that animations with narrations plus text were more effective than animations with text only for factual learning performance. The final significant finding was that participants assigned to text only conditions, regardless of graphic type, spent more additional time waiting at the end of an instructional segment than participants who received narrations alone. The following is a discussion of these findings. The discussion then addresses some of the methodological issues for this research and concludes with considerations for future research.

Differences between Animations with Narrations and Static Graphics with Narrations

Although there was no difference on factual learning, participants who received animations with narrations performed significantly better on transfer measures than participants who received static graphics with narrations. At first one might conclude that this finding supports predictions, but this support must be qualified. The lack of significant main effects for graphic type and description modality or interaction on transfer scores suggests that there should be no differences between animations and static graphics with narrations. More important, the superiority of transfer for animations with narrations to static graphics with narrations, shown in figure 4, consists of two parts. For animations, the narrations mean is *higher* than the means for text only and narrations plus text. For static graphics, the narrations mean is *lower* than the means for text only and narrations plus text. The significant difference is a combination of an improvement for animations and degradation for static graphics with narrations relative to the other description modalities. However, the non-significant simple main effects for description modality at the two levels of graphic type indicate an absence of evidence for differences between narrations plus text, narrations only or text only for animations and also for static graphics. In figure 4, both the "improvement" that one observes for animations with narrations and "degradation" observed for static graphics with narrations exist in the midst of other evidence that they should be attributed to random variation.

In addition to the contradictory statistical evidence, there are logical difficulties with the significant difference between animations with narrations and static graphics with narrations on transfer. The improvement in animations with narrations relative to animations with text only and animations with narrations plus text makes sense. But the decrease in performance for static graphics with narrations relative to static graphics with text only and static graphics with narrations plus text does not make sense. Without the decrease for static graphics relative to text and narrations plus text, the comparison of animated vs. static graphics with narrations would not be significant. Unless a plausible explanation for the decrease in transfer with static graphics and narrations can be found, the significant finding should be provisional.

Differences between Description Modalities for Animations

On factual learning measures, participants who received animations with narrations plus text performed better than participants who received animations with text alone. This difference was not obtained for participants assigned to the equivalent static graphics condition suggesting that some aspect of the animations influenced participants' processing and comprehension of the lesson materials that the static graphics did not. The finding that animations with narrations plus text are better than animations with text alone goes against the claim (Mayer et al., 2001b) that animations with narrations are the most effective combination for enhancing learning through animations. According to Mayer et al., animations with narrations plus text should hinder performance since visual working memory is overloaded by the animations and text (Mayer et al., 2001b).

One possible explanation for the benefit of narrations plus text versus text alone is the interplay between the presentation of multiple media and the inclusion of learner-pacing in the form of the "proceed" button that appeared at the end of each segment. This speculation is supported by the sample data collected, but not by statistical tests. The speculations are offered because of the dramatic and potentially important contrast between the results reported by the Mayer studies that used pure animations and the results found here when learner-control was provided.

In the case of animations with narrations plus text, participants might have watched the animation and listened to the explanation, and *then* read the on-screen text when the animation and narrations stopped. This was possible because the text remained on screen until the "proceed" button was clicked to advance to the next segment. Thus, learners might have processed the descriptive material twice: once by listening to the narrations while ignoring the text and then by reading the text once the narrations were complete. Had the learner-pacing control not been available, participants would not have been able to attend to the on screen text as easily because successive segments would appear immediately after the previous segment ended. If participants did, in fact, process the material in this way, they may have improved their recognition performance.

By looking at the percentage of additional time participants spent waiting at the end of lesson segments, one can make some plausible speculations that add support to the learner-pacing explanation for performance on factual learning scores. Given that the entire lesson took approximately three minutes, each segment of the lesson lasted approximately 11.3 seconds (180 seconds divided by 16 lesson segments). When receiving animations with narrations plus text, participants spent a mean of 3.3 seconds per segment waiting at the end of segments versus 1.9 seconds per segment for those who received animations with narrations alone. Therefore, participants who received animations with narrations plus text spent 30% more time waiting at the end of segments compared to 16% more time for participants who received animations with narrations only. Although this difference was not significant, participants who had on-screen text with narrations spent almost twice as much time waiting as those who had narrations only. It makes sense that participants who had on-screen text waited longer because, in the animations with narrations condition, there was no on-screen text to read after the animation finished. That participants spent longer waiting when presented with narrations plus text suggests they used this time to read or re-read the on-screen text, thereby enhancing factual learning performance. It is important to note that the preceding speculation is not warranted by the statistical results. Given the statistics, one cannot make definitive statements about participants who received narrations plus text taking longer than participants who received narrations alone since these conditions did not differ statistically.

Participants assigned to animations with text only, who spent 40% more time waiting at the end of segments, did not perform as well as those assigned to animations with narrations plus text. The difference in additional time between participants who received text only and narrations only was not significant, and it was not significantly different from participants who received narrations plus text. Learners who received text alone spent a mean of 40 seconds longer at the end of lesson segments than those who received narrations alone but did not perform any better on factual learning scores, and they spent 20 seconds longer than participants in the narrations plus text condition and performed worse. One possible explanation for the lack of performance gains for participants who received animations with text only is that these participants did not have the advantage of the double presentation of the description modalities; they could only read the on-screen text and did not hear the narrations.

One may wonder how participants might process text while static images are presented. One possibility is that, with the simple concepts conveyed by arrows in this tutorial, the static images did not occupy participants for long. Participants could look at the arrows, quickly process the information in the arrows, and then turn attention to reading the text. On the other hand, with animations, participants had to watch as the action transpired, not knowing what would happen. While animations conflicted with text reading until the action completed, static arrows might have had a smaller detrimental effect on text reading. If this is true, one might expect performance on static graphics with text to be better than performance on animations with text, especially with learner control. As shown on the right side of figure 3, there is a trend in that direction.

The issue of pacing is central to the findings in this thesis. Mayer's research has, with one exception, used animated lessons that played through from beginning to end with no opportunity for students to control pace. What benefits can learners derive from animated lessons that play through with no opportunity to pause, especially if learners are presented with both narrations and text? If learners miss a detail in the lesson, they are forced to watch the rest of the tutorial before being able to return to the portion they missed or they have no opportunity to review at all. In a non-interactive environment such as television or movies, pure animations, which are continuous and do not allow for learner control, can be justified. In an interactive environment such as computer-based training, CD and web-based presentations, it is not clear what the advantages of pure animations are.

One study from Mayer (2001a) used learner-pacing of lesson materials but that study did not compare learner-paced with pure uninterrupted animations in a straightforward way. Participants in that study were presented the tutorials twice, such that they received combinations of learner-paced or pure animations presented successively. Results were that presenting learners with a learner-paced animation before presenting them with a pure uninterrupted animation or presenting two consecutive learner-paced animations resulted in improved transfer performance over participants who received a pure uninterrupted animation before a learner-paced animation or two consecutive pure uninterrupted animations.

Mayer's (2001) learner-pacing study does not contribute to the findings from this study for several reasons. First, Mayer's lesson contained narrations only but not narrations plus text where one would expect to find advantages of learner pacing. Second, for whatever reason, Mayer presented lesson materials to participants twice, making it difficult to compare Mayer's results to results from this research because participants in this study only saw lesson materials once.

Based on the assumption that it may be detrimental for learning to force learners to watch lessons with no opportunity to control pauses, it follows that some of Mayer's recommendations about animations may be inappropriate because they were all based on pure animations. Specifically, the finding from this research that animations with narrations *plus* text resulted in the best factual learning performance rather than the worst suggests that when learners are given an opportunity to process information by allowing pauses they are at an advantage if they have access to narrations and text. Mayer has summarized the findings of his research in seven principles of multimedia learning (Mayer, 2002). With the exception of the Redundancy Principle, which says that text should not be presented with narrations, most of Mayer's recommendations seem to apply whether graphic materials are learner–paced or automatically paced. Future studies should address whether Mayer's recommendations are valid when learner-paced lesson materials are used.

Related to pacing is the notion of interactivity discussed by Tversky et al. (Tversky et al., 2002). Tversky et al. suggest that interactivity, *not* the multimedia per se may be an important facilitating factor for learning when multimedia is used. One of the main criticisms Tversky et al. levied against studies exploring animations was the inclusion of interactive elements in the animated conditions that did not appear in the static graphics conditions. In this study, interactivity was represented by a "proceed button" that allowed learner-pacing of the lesson materials. While the experimental materials in this study contained an element of interactivity with the inclusion of the "proceed" button, both the animated and static graphics conditions made use of the same "proceed" button and all other lesson content was equivalent. Thus, this thesis was not about pure animations.

While learner-pacing was a relatively minor form of interactivity in this study, there are other important ways that interactivity could be used in lesson materials. Examples of more complex interactivity include, but are not limited to, hyperlinks that take learners to more detailed information and provision for manipulation of virtual objects. There are many potential benefits of these more complex types of interactivity versus pure animations. For example, imagine a lesson about car engines for mechanics where the learner may be able to select a specific part of the engine, "pull" the part out virtually and rotate it. The learner may then click on the object, which starts a learner-paced animation with narrations that allows the learner to explore the part at a pace that suits the individual's learning style. In this interactive scenario, the learner not only determines what is viewed and the order the lesson materials are viewed but also at what pace. These could have advantages over an animation that plays through from beginning to end with no pause.

Methodological Issues

One issue in this research was the nature by which the experimental materials were constructed in an attempt to replicate the tutorials used by Mayer. Attempts were made to obtain the original lesson materials used by Mayer, but due to unwillingness of the creator of the materials to make the lessons available, I was forced to recreate the lessons based on descriptions from research articles. The drawback is that while some content, such as quality of graphics and wording of the text and narrations seemed easy to duplicate, there may have been some details that could not be identified through research articles. That an exact replication of the lesson materials could not be made means that there may be unknown factors that lead to differences in findings between Mayer's work and this research.

Another issue, related to the replication of the lesson materials, is the use of multiple-choice questions rather than free recall to measure performance. In Mayer's research, post test questions involved hand written, open-ended recall. Recall questions force learners to do two things -- generate alternatives and then select from them. On the other hand, recognition provides the alternatives so the task is to select among the alternatives (Driscoll, 2000). In this study, participants had to recognize the one correct response out of four possible responses. Recognition is usually easier than recall because the first part of the process, provide the alternatives, is already done (Driscoll, 2000). It is possible that factors such as interactivity, graphic type and description modality may have different effects when performance is measured by multiple-choice questions and recall for factual learning and transfer questions. It is particularly challenging to convert transfer questions into multiple-choice questions since transfer questions are designed to assess whether learners can generate solutions to unique problems. Presenting transfer questions in a multiple-choice format merely presents the learner with a choice between a number of unique answers rather than having them create their own unique solutions. Generally, transfer questions are difficult to generate and determining the correct response to a transfer question is difficult as evidenced by the criticism of Mayer's correct transfer responses discussed earlier. Of the 19 questions participants answered in this study only three were transfer questions. To further explore differences in performance between animations and static graphics, more and better transfer questions should be included. As evidenced by the reliability measures obtained for the factual learning and transfer questions, it is challenging to generate multiple choice-questions that measure specific types of learning. That ceiling or floor effects were not observed in the performance results is not sufficient to make conclusions that questions were properly formed. To make statements about factual and transfer learning, future studies must create reliable measures of these learning types regardless of whether they are presented as multiple-choice, free recall or transfer questions.

Another issue, related to the multiple-choice questions, is that they were presented in the approximate chronological order of the lesson materials. Specifically, questions 1 through 5 dealt with ideas presented at the beginning of the lesson while questions 14, 15, and 16 dealt with ideas presented at the end of the lesson. It may be that results might have been different had the questions been in a less meaningful order.

A particularly perplexing aspect of all the research performed by Mayer and his associates is the seemingly inconsistent pattern of results that apply sometimes to factual learning, sometimes to transfer and sometimes to both. What are the mechanisms that account for these inconsistencies? In this thesis, how could we explain why animation might be better than static graphics for transfer, but not for factual learning? One possible explanation stems from the lack of reliability of the measures. It may be the case that the measures Mayer has been using are not reliable either. In addition, Mayer often refers to transfer learning as "deep understanding" (Mayer & Chandler, 2001a) of lesson materials. That transfer is considered deep understanding implies that factual learning measures represent a more superficial understanding of lesson materials. If a learner has successfully processed the learning materials at a deep enough level that performance on transfer problems are enhanced, why should performance on factual learning measures not also be enhanced? One would expect that the performance on the factual learning measures would be at least as good as or better than transfer performance, but this has not always been the case in the Mayer studies since benefits of animations are often only realized for transfer scores.

Finally, a discussion of the lightning tutorial itself is warranted. For the purposes of this study it was advantageous to attempt replication of the tutorial used by Mayer. However, continued use of this tutorial will reduce our ability to make generalizations about the effects of animations and description modalities. Furthermore the topic of lightning formation itself, presented at the difficulty level in this study, may not fully demonstrate the potential advantages for animations over static graphics. In this study the formation of lightning was presented at a very general level, but it is highly likely that a meteorologist would find the material too basic.

Future Directions

Many questions arise from this study. Future research should focus on the following issues to further our understanding of the interplay between graphic type and description modality.

A deeper exploration into the effects of pacing should be conducted. Specifically, comparisons should be made between combinations of description modalities with both learner-paced and pure animation lessons. If it is the case that learner-paced animations allow learners to effectively use greater combinations of media, such as text plus narrations, then the use of pure animations is of questionable value. It is also of interest to confirm that Mayer's recommendations all apply with learner-paced animations. Future studies should also explore how other types of interactivity, such as allowing manipulation of on-screen graphics or including hyperlinks that lead to further detail, can be used to support animations and description modalities to enhance factual learning and transfer performance.

The finding that performance on transfer questions decreased for static graphics with narrations relative to static graphics with text plus narrations or text only is not expected from other research performed on narration, text and text plus narrations. A study of static graphics while varying the three description modalities for factual learning and transfer questions would be one effective means to resolve the issue. Alternatively, a replication of this thesis research would also be a means of confirming that static graphics with narrations reduce transfer performance compared to static graphics with text plus narrations or text only.

Mayer and his colleagues have always used free recall combined with transfer to explore the effects of animations and narrations. Future research should compare the differences in performance for recall and recognition when using combinations of graphic type and description modalities as lesson materials. It may be that Mayer's results are only relevant for free recall and transfer questions and not for multiple-choice recognition questions.

Given Mayer's focus on transfer questions being the best indicators of deep understanding of lesson materials, future studies should make explicit comparisons between performance on recognition or recall questions versus transfer questions. If benefits for animations and description modalities are seen primarily when measured by transfer questions, there are implications for the design of testing materials. Since it is difficult to generate effective transfer questions and solutions, it will be problematic to consistently create robust and varied test materials. Another implication, for students, is that they may not know how to effectively answer a transfer-type question. These are important considerations for the design of animated lesson materials if benefits for them are only or primarily realized on transfer scores.

Research that aids in understanding when animation will help transfer and when it will help factual learning, and why such seemingly arbitrary results are obtained will be useful.

However, one must be certain to use reliable measures of factual learning and transfer before pursuing further research on this topic.

Future studies should also explore lesson topics other than the formation of lightning, or should at least present the lightning materials at a more complex level than was presented in this study. One of the major difficulties in creating the experimental materials for this study was the selection of lesson topic. The topic should be presented at a level complex enough so that ceiling effects are not apparent in participant performance. There were some initial reservations about the use of the lightning tutorial as there was concern that the lesson would be too simplistic, however the final results do not indicate ceiling effects in participant performance.

Wong (1993), Hutcheson (1997) and I all searched extensively for animations that would be appropriate for comparisons of animations and still graphics. In almost all cases, after considering the animation and what the appropriate comparable still images would be, it was clear that the still images could be as effective as the animations.

Related to the selection of topic is the appropriateness of the topic. It was agreed that whatever topic was chosen should, as part of the fundamental nature of the topic, deal with motion, time, change of state, demonstrate a process or to aid in demonstrating an abstract concept through metaphor. It does not make sense to create an animation for animations sake.

Reference List

Baek, Y. K. & Layne, B. H. (1988). Color, Graphics, and Animation in A Computer-Assisted-Learning Tutorial Lesson. *Journal of Computer-Based Instruction, 15,* 131-135.

Driscoll, M. (2000). *Psychology of Learning for Instruction.* Toronto: Allyn & Bacon.

Hegarty, M., Quilici, J., Narayanan, N. H., Holmquist, S., & Moreno, R. (1999). Multimedia Instruction: Lessons from Evaluation of a Theory-based Design. *Journal of Educational Multimedia and Hypermedia, 8,* 119-150.

Hutcheson, T. (1997). *The Effectiveness of Animation and Narration in Computer-Based Instruction.* Unpublished master's thesis, Carleton University.

Large, A., Beheshti, J., Breuleux, A., & Renaud, A. (1996). Effect of animation in enhancing descriptive and procedural texts in a multimedia learning environment. *Journal of the American Society for Information Science, 47,* 437-448.

Macromedia (2002). Director MX [Computer software].

Mayer, R. E. (2002). Multimedia learning. *Psychology of Learning and Motivation: Advances in Research and Theory, 41,* 85-139.

Mayer, R. E. & Anderson, R. B. (1991). Animations Need Narrations - An Experimental Test of A Dual- Coding Hypothesis. *Journal of Educational Psychology, 83,* 484-490.

Mayer, R. E., Bove, W., Bryman, A., Mars, R., & Tapangco, L. (1996). When less is more: Meaningful learning from visual and verbal summaries of science textbook lessons. *Journal of Educational Psychology, 88,* 64-73.

Mayer, R. E. & Chandler, P. (2001a). When learning is just a click away: Does simple user interaction foster deeper understanding of multimedia messages? *Journal of Educational Psychology, 93,* 390-397.

Mayer, R. E., Heiser, J., & Lonn, S. (2001b). Cognitive constraints on multimedia learning: When presenting more material results in less understanding. *Journal of Educational Psychology, 93,* 187-198.

Mayer, R. E. & Moreno, R. (1998). A Split-attention effect in multimedia learning: Evidence for dual processing systems in working memory. *Journal of Educational Psychology, 90,* 312-320.

Mayer, R. E. & Moreno, R. (2002). Animation as an aid to multimedia learning. *Educational Psychology Review, 14,* 87-99.

Mayer, R. E., Moreno, R., Boire, M., & Vagge, S. (1999). Maximizing constructivist learning from multimedia communications by minimizing cognitive load. *Journal of Educational Psychology, 91,* 638-643.

Mayer, R. E. & Sims, V. K. (1994). For Whom Is A Picture Worth 1000 Words - Extensions of A Dual- Coding Theory of Multimedia Learning. *Journal of Educational Psychology, 86,* 389-401.

Moreno, R. & Mayer, R. E. (1999). Cognitive principles of multimedia learning: The role of modality and contiguity. *Journal of Educational Psychology, 91,* 358-368.

Moreno, R. & Mayer, R. E. (2000). A coherence effect in multimedia learning: The case for minimizing irrelevant sounds in the design of multimedia instructional messages. *Journal of Educational Psychology, 92,* 117-125.

Park, O. C. & Gittelman, S. S. (1992). Selective Use of Animation and Feedback in Computer-Based Instruction. *Etr&D-Educational Technology Research and Development, 40,* 27-38.

Rieber, L. P. (1990). Using Computer Animated Graphics in Science Instruction with Children. *Journal of Educational Psychology, 82,* 135-140.

Rieber, L. P. (1991). Animation, Incidental-Learning, and Continuing Motivation. *Journal of Educational Psychology, 83,* 318-328.

Schnotz, W., Bockheler, J., & Grzondziel, H. (1999). Individual and co-operative learning with interactive animated pictures. *European Journal of Psychology of Education, 14,* 245-265.

Tversky, B., Morrison, J. B., & Betrancourt, M. (2002). Animation: can it facilitate? *International Journal of Human-Computer Studies, 57,* 247-262.

Wong, A. (1993). *The Use of Animation in Computer Assisted Instruction.* Unpublished doctoral dissertation, Carleton University.

Appendix A

Descriptions for each tutorial segment

1. Cool moist air moves over a warmer surface and becomes heated
2. Warmed moist air near the earth's surface rises rapidly
3. As the air in this updraft cools, water vapour condenses into water droplets and forms a cloud
4. The cloud's top extends above the freezing level, so the upper portion of the cloud is composed of tiny ice crystals
5. Eventually, the water droplets and ice crystals become too large to be suspended by the updrafts
6. As raindrops and ice crystals fall through the cloud, they drag some of the air in the cloud downward, producing downdrafts
7. When downdrafts strike the ground, they spread out in al directions, producing the gusts of cool wind people feel just before the start of the rain
8. Within the cloud, the rising and falling air currents cause electrical charges to build
9. The charge results from the collision of the cloud's rising water droplets against heavier, falling pieces of ice
10. The negatively charged particles fall to the bottom of the cloud, and most of the positively charged particles rise to the top
11. A stepped leader of negative charges moves downward in a series of steps. It nears the ground
12. A positively charged leader travels up from such objects as trees and buildings
13. The two leaders generally meet about 165-feet above the ground

14. Negatively charged particles then rush from the cloud to the ground along the path created by the leaders. It is not very bright
15. As the leader stroke nears the ground, it induces an opposite charge, so positively charged particles from the ground rush upward along the same path
16. This upward motion of the current is the return stroke. It produces the bright light that people notice as a flash of lightning

Appendix B

Knowledge of meteorology questionnaire and Self Rating Scale

Participants were asked to place checks next to all of the following that applied:

"I know what a cold front is"

"I can distinguish between cumulous and nimbus clouds"

"I know what a low pressure system is"

"I can describe how lightning forms"

"I know what this symbol means" (symbol for warm front)

"I know what this symbol means" (symbol for cold front)

"I can explain what makes the wind blow."

Self assessment was measured using a five-point scale that asked participants to rate their knowledge of meteorology by placing a check next to one of:

"very little"

"between very little and average"

"average"

"between average and very much"

"very much."

Appendix C

19 Recognition Questions Broken into Learning and Transfer

<u>Multiple-choice (Learning)</u>

1. What type of air moves over the Earth's surface at the start of the process that results in lightning?
    a. Moist and hot
    b. Cool and moist
    c. Humid and warm
    d. Dry and hot
2. During the formation of lightning the temperature of the Earth's surface causes the air to:
    a. Condense
    b. Humidify
    c. Rise
    d. Fall
3. What happens to water vapour before forming lightning storm clouds?
    a. It contracts
    b. It cools
    c. It expands
    d. It condenses
4. For clouds to contribute to the formation of lightning they must:
    a. Expand in size
    b. Decrease their mass
    c. Rise above the freezing level

d. Sink towards the Earth

5. Many things form inside clouds as part of the formation of lightning, one of them includes:
   a. Ice crystals
   b. Mini-cyclones
   c. Hail stones
   d. Water droplets
6. Downdrafts that occur before lightning storms are the result of?
   a. Large ice crystals melting and evaporating
   b. Large ice crystals colliding and forming other clouds
   c. Large ice crystals and water droplets falling from the cloud
   d. The difference in temperature between cool clouds and the warm land
7. The gusts of cool wind people feel before seeing lightning indicate what?
   a. That a low pressure system has settled over the area
   b. That a high pressure system has settled over the area
   c. That converging air masses are causing the air temperature to drop
   d. That rain is about to fall
8. Rising and falling air currents cause:
   a. Rising water droplets to collide with falling pieces of ice resulting in electrical charges within the cloud
   b. Rising dust particles to collide with falling pieces of ice resulting in electrical charges in the cloud
   c. Rising water droplets to collide with falling dust particles resulting in electrical charges within the cloud

d. Rising dust particles to collide with falling water droplets resulting in electrical charges within the cloud

9. What happens to negative and positive charges within the cloud during the formation of lightning?
   a. Negative and positive charges collide and discharge
   b. Negative and positive charges rise to the top of the cloud
   c. Negative charges fall to the bottom of the cloud and positive charges rise to the top
   d. Positive charges fall to the bottom of the cloud and negative charges rise to the top
10. A stepped leader of charged particles moves downwards from the cloud; what charge does the leader have?
   a. Positive
   b. Negative
   c. Neutral
   d. Static
11. A stepped leader of charged particles moves upwards from the ground; what charge does the leader have?
   a. Positive
   b. Negative
   c. Neutral
   d. Static
12. Stepped leaders moving downwards from clouds and upwards from the ground meet at what distance from the Earth's surface?
   a. 156 ft

b. 163 ft

c. 164 ft

d. 165 ft

13. During the formation of lightning charged particles rush from the cloud to the ground along the path of the leaders; what charge do the particles have?

    a. Positive

    b. Negative

    c. Neutral

    d. Static

14. The charged particles that rush from the cloud to the ground along the path of the leaders create light. What do we know about the light that is created?

    a. It is perceived as a bolt of lightning

    b. It is bright

    c. It is not visible at all

    d. It is not very bright

15. During the formation of lightning what happens when the downward moving stepped leader nears the ground?

    a. It attracts to tall objects such as buildings and trees

    b. It induces a parallel charge so positively and negatively charged particles from the ground rush upward along the same path

    c. It induces an opposite charge so positively charged particles from the ground rush upward along the same path

    d. It strikes the ground and is perceived as a bolt of lightning

16. The "return stroke" refers to what?

   a. The bright light that is perceived as a bolt of lightning
   b. The negative charges that rush upwards towards the sky
   c. The combination of negative and positive charges that rush upwards towards the sky
   d. The negative charges that rush downwards towards the ground

<u>Multiple-choice (transfer)</u>

17. Which of the following could conceivably decrease the intensity of lightning?

   a. Increase positive charges in the air
   b. Remove negative charges from the ground
   c. Increase the temperature of the air near the Earth's surface
   d. Remove positive charges from the ground

18. Why might you see clouds in the sky but no lightning?

   a. The clouds are part of a cold front and will never form lightning
   b. The clouds are part of a warm front and will never form lightning
   c. The tops of the clouds are not yet above the freezing point
   d. The air near the Earth's surface was not strong enough to create charged particles in the cloud

19. Which of the following statements is true?

   a. Cool air produces an abundance of downdrafts that promote the formation of lightning
   b. Cool air is a better conductor for charged particles to pass from the ground to the cloud

c. Warm air is a better conductor for charged particles to pass from the ground to the cloud

d. The air must be cooler than the ground for lightning to form

Appendix D

Debriefing

Thank-you for agreeing to participate in this study. This is an informed consent form that states what will be involved in participating in this study. The purpose of an informed consent form is to explain to you, as a participant, what is involved in participating in the study so that you can make an informed decision regarding whether or not you wish to participate.

Today you will be looking at a computer-based tutorial on the formation of lightning. You will be required to answer some questions regarding your knowledge of meteorology, view the tutorial which lasts approximately three minutes, and then answer a series of questions about the tutorial. These questions should take approximately fifteen minutes to complete. You will be interacting with a computer via a mouse by advancing the tutorial by clicking on buttons or answering multiple-choice questions.

The purpose of this study is to determine how computer based tutorials can be effective in aiding recall of facts.

While there are no aspects to this study that should cause you any discomfort, please be assured that if at any time you feel you are unable or do not wish to complete the session you may stop. If there are questions you find objectionable or do not wish to answer you are not required to answer them. You will still receive credit for the time you have participated.

Your identity will remain completely anonymous throughout the duration of this research; rather than using actual names, you will be assigned a unique number that will identify you in any written reports or discussions about this study. Data from this study will only be shared with the research personnel involved.

I have read the informed consent form and agree to participate in this study:

____________________ ____________________

Name Date

Appendix E

Table E1

ANOVA Summary Table for Graphic Type and Description Modality for all 19 Questions

Source	Sum of Squares	df	Mean Square	F	p
Graphic Type	52.02	1	52.02	0.30	.58
Description Modality	820.56	2	410.20	2.40	.10
Graphic Type x Description Modality	340.41	2	170.21	0.99	.37
Error	14363.80	84	171.00		
Total	15576.79	89			

Appendix F

Table F1

ANOVA Summary Table for Graphic Type and Description Modality for Prior Knowledge

Source	Sum of Squares	df	Mean Square	F	p
Graphic Type	0.18	1	0.18	0.07	.80
Description Modality	0.82	2	0.41	0.15	.86
Graphic Type x Description Modality	4.96	2	2.48	0.92	.40
Error	225.60	84	2.69		
Total	231.56	89			

Appendix G

Table G1

Means and Standard Deviations (in parentheses) for Graphic Type and Description Modality for Prior Knowledge

	Narrations + text	Narrations	Text	Total
Animations	2.13	2.60	1.80	2.18
	(1.92)	(1.59)	(1.15)	(1.55)
Static graphics	2.33	2.07	2.40	2.27
	(1.80)	(1.58)	(1.68)	(1.69)
Total	2.22	2.33	2.10	2.22
	(1.61)	(1.58)	(1.41)	(1.57)

Appendix H

Graph for Graphic Type, Description Modality and Prior Knowledge

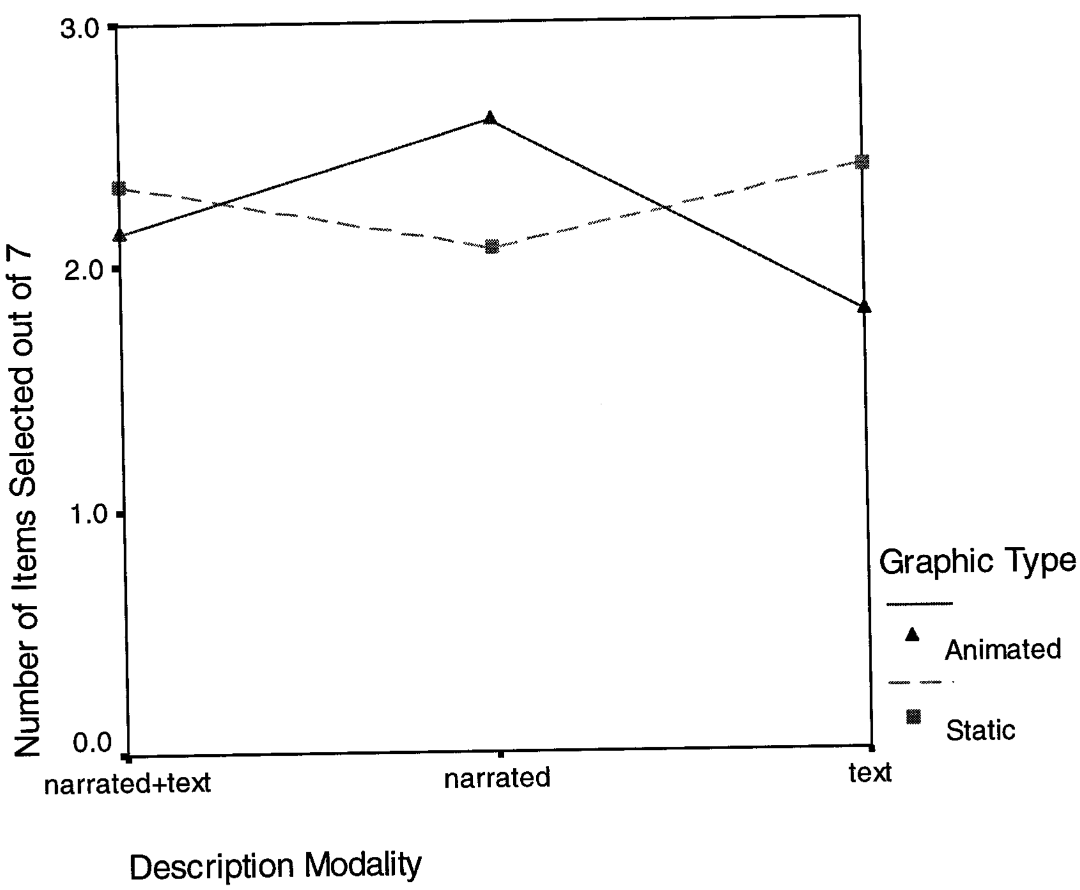

Figure H1. Mean responses to prior knowledge questions

Appendix I

Table I1

MANOVA Summary Table for Graphic Type and Description Modality for Learning, Transfer and Additional Time

Effect	Statistic	Value	F	Hypothesis df	Error df	p
Graphic Type	Pillai's Trace	0.07	1.99	3.00	82.00	.12
	Wilks' Lambda	0.93	1.99	3.00	82.00	.12
	Hotelling's Trace	0.07	1.99	3.00	82.00	.12
	Roy's Largest Root	0.07	1.99	3.00	82.00	.12
Description Modality	Pillai's Trace	0.24	3.79	6.00	166.00	.001
	Wilks' Lambda	0.77	3.81	6.00	164.00	.001
	Hotelling's Trace	0.28	3.82	6.00	162.00	.001
	Roy's Largest Root	0.21	5.93	3.00	83.00	.001
Graphic Type x Description Modality	Pillai's Trace	0.09	1.34	6.00	166.00	.24
	Wilks' Lambda	0.91	1.34)	6.00	164.00	.24
	Hotelling's Trace	0.10	1.35	6.00	162.00	.24
	Roy's Largest Root	0.09	2.46	3.00	83.00	.07

Appendix J

Table J1

ANOVA Summary Table for Graphic Type and Description Modality for Learning Questions

Source	Sum of Squares	df	Mean Square	F	p
Graphic Type	6.94	1	6.94	1.36	.25
Description Modality	30.82	2	15.41	3.01	.05
Graphic Type x Description Modality	7.22	2	3.61	0.71	.50
Error	429.33	84	5.11		
Total	474.32	89			

Appendix K

Table K1

Means and Standard Deviations (in parentheses) for Graphic Type, Description Modality and Learning Performance

	Narrations + text	Narrations	Text	Total
Animations	83.33	76.25	70.42	76.67
	(12.87)	(16.74)	(18.97)	(16.19)
Static graphics	83.33	78.33	78.75	80.14
	(7.34)	(13.54)	(12.45)	(11.11)
Total	83.33	77.29	74.58	78.40
	(10.29)	(15.00)	(16.33)	(13.78)

Appendix L

Table L1

ANOVA Summary Table for Graphic Type and Description Modality for Transfer Questions

Source	Sum of Squares	df	Mean Square	F	p
Graphic Type	1.60	1	1.60	3.18	.08
Description Modality	0.02	2	0.01	0.02	.98
Graphic Type x Description Modality	1.40	2	0.70	1.39	.25
Error	42.27	84	0.50		
Total	374.00	90			
Corrected Total	45.29	89			

Appendix M

Table M1

Means and Standard Deviations (in parentheses) for Graphic Type, Description Modality and Transfer Performance

	Narrations + text	Narrations	Text	Total
Animations	66.67	73.33	64.44	68.15
	(21.82)	(22.54)	(19.79)	(21.27)
Static graphics	60.00	53.33	64.44	59.26
	(28.73)	(24.56)	(23.46)	(25.51)
Total	63.33	63.33	64.44	63.70
	(25.29)	(25.29)	(21.32)	(23.74)

Appendix N

Table N1

ANOVA Summary Table for Graphic Type and Description Modality for Total Additional Time

Source	Sum of Squares	df	Mean Square	F	Sig.
Graphic Type	839.26	1	839.26	1.16	.28
Description Modality	12486.50	2	6243.25	8.67	.000
Graphic Type x Description Modality	2547.37	2	1273.68	1.77	.18
Error	60510.46	84	720.36		
Total	76383.59	89			

Appendix O

Table O1

Means and Standard Deviations (in parentheses) for Graphic Type, Description Modality and Total Additional Time

	Narrations + text	Narrations	Text	Total
Animations	51.98	29.69	71.53	51.06
	(34.86)	(17.06)	(51.06)	(32.97)
Static graphics	45.96	36.57	52.34	44.96
	(24.19)	(20.08)	(29.26)	(25.10)
Total	48.97	33.13	61.94	48.01
	(29.64)	(18.64)	(31.26)	(29.30)