I, Andrew Curran, hereby submit this original work as part of the requirements for
the degree of Doctor of Education in Curriculum & Instruction.

It is entitled:
The Effect of Adding Relevant Music and Sound Effects to an Audio-Only
Narration:
A Three-Treatment Application of Mayer's Coherence Principle

Student's name: Andrew Curran

This work and its defense approved by:

Committee chair: Maya Israel, PhD
Committee member: James Basham, PhD
Committee member: John Owens, PhD
Committee member: Kay Seo, PhD
The Effect of Adding Relevant Music and Sound Effects to an Audio-Only Narration:

A Three-Treatment Application of Mayer’s Coherence Principle

A dissertation submitted to the

Graduate School

of the University of Cincinnati

in partial fulfillment of the

requirements for the degree of

Doctor of Education

in the Department of Instructional Design and Technology

of the School of Education

of the College of Education, Criminal Justice, and Human Services

by

Andy Curran

M.Ed., University of Cincinnati

B.S.C., Ohio University

July 2012

Committee Chair: Maya Israel, Ph.D.
Abstract

Sound effects and music can be used to complement instructional narrations in recorded media presentations. College students \((n = 143)\) participated in a study examining the effects of music and sound effects on learning outcomes when compared to voice-only narration. Four groups of participants listened to a recorded short story and answered questions about the story. The control group’s story was a voice-only narration. The three treatment groups’ story was enhanced with one of the following: sound effects, music, or a combination of sound effects and music. Means for each group were not significantly different, and the difference of means between and within each of the groups was also found not to be significant. The results indicated that the presence sound effects and music from an audio-only presentation did not significantly increase or decrease learning.

*Keywords*: sound effects, music, audio, coherence principle, arousal theory, multimedia learning
Acknowledgments

This five-year academic journey has been a fulfilling experience, but it would not have been possible without assistance. First, I am grateful that the University of Cincinnati gave me the opportunity to pursue this dream because of the employee Tuition Remission benefit. I appreciated the encouragement of Greg Sojka, Dean of UC-Clermont College, his predecessor, Jim McDonough, and my current and former department chair, Jeff Bauer and Tracey Hawkins, respectively, for allowing me to take on extra credit hours to fulfill program requirements.

The faculty at UC taught me more about educational research than I thought I would ever know. The most valuable piece of knowledge that I took from them is that educational research has a human side that is not present in some other disciplines, so there is a need to approach research carefully. In particular, I would like to thank my four dissertation committee members: Drs. Maya Israel (Chair) and Kay Seo from the UC College of Education, Criminal Justice, and Human Services, Dr. John Owens from UC’s College-Conservatory of Music, and Dr. Jamie Basham from the University of Kansas. In the classroom, Dr. Seo’s seminars helped me develop my writing and research skills, and her studio courses gave me greater insight into the practical application of pedagogy in instructional technology. Dr. Owens introduced me to a body of academic audio research that formed a major part of this dissertation. Dr. Basham was my original advisor and committee chair, and he pointed me in the right direction early in my studies, when my research agenda had no focus. His guidance enabled me to sharpen my academic writing and research skills, which enabled me to get my Phase I paper published in a conference proceeding. Dr. Israel assumed the role of my advisor and committee chair in 2011, when Dr. Basham went to KU. Changes in dissertation committee chairs can be tenuous, but Dr. Israel quickly eased my anxieties by encouraging me to continue on the same research track. She
was the key person who helped me shape this dissertation by making sure I included details about the methods and instrumentation, statistics and data verifications, and the organization of the content. I could never thank the committee enough for helping me get to this point. Back in the mid-1990s, I earned my Master of Education in Instructional Design and Technology at UC. My advisor, Dr. Randy Nichols, who is no longer with UC, was the person who first showed me how to navigate the rigors of graduate school and how to tackle complex research projects.

I would also like to acknowledge the support of my fellow graduate students, especially my cohorts, Aimee DeNoyelles and Dana Tindall. Early on, when all of us were trying to figure out what we were going to do, we helped each other by sharing our knowledge and commiserating, usually over a beer at Buffalo Wild Wings or dinner at Panera.

If I didn’t have a 25-year radio career, I wouldn’t have been interested in audio research. There were some very good radio mentors who taught me a lot about how to properly produce audio. People like Ty Richards, Bob Cooper, the late Rusty Walker, Randy Lane, Greg Pencheff, Howard Galvin, and others were instrumental in helping me have a successful run in radio.

Finally, many thanks to my family. It’s impossible to get into graduate school without first earning a Bachelor’s degree. If not for the financial generosity and moral support of my parents, Pat and Joe, I would not have been a member the Ohio University Class of 1977, so it is likely that I wouldn’t have written this dissertation. My younger sisters, Pat and Nancy, and I were first generation college graduates in our family. My mom didn’t go to college because, unfortunately, a lot of women were discouraged from doing so in 1947. My dad attended Cornell University for two years, but he had to withdraw for personal reasons. I know that he wished he could have gotten his degree, but his family came first and he had to support and raise us. He passed away in 2010, and I regret that he wasn’t able to see me complete this degree. I am
confident that he would be proud of me. My daughters, Carolina, Monica, and Melinda, have been wonderful throughout this process, although they have trouble believing that their “goofy” dad might become known as Dr. Curran. But I wouldn’t be the first Dr. Curran in our family. That honor belongs to my lovely and talented wife, Chris, who earned her Ph.D. in Environmental Health at UC in 2007, and has just completed her fourth year as an Assistant Professor of Physiology at Northern Kentucky University. She is the smartest person I have ever known personally, and she has been a great help with her encouragement and insight into academic research. She was the one who inspired me to pursue this degree.

Although I consider myself a “lifelong learner”, this is the terminus of my formal education, which started in kindergarten at Public School 128 in Brooklyn, NY, in 1960, with stops at St. Mary’s Elementary School in Brooklyn, Candlewood Junior HS on Long Island, Randolph HS in northern New Jersey, Worthington HS in suburban Columbus, and, as I mentioned, OU and UC. The ride was interesting, and not always smooth. But it was worth it, thanks to the teachers and professors, most of them unnamed, at those institutions of learning. This might be the end of my formal education, but I hope it is just the beginning of my professional research agenda. I subscribe to the philosophy of singer John Mayer, who sang these lyrics in the song, “No Such Thing”: “I’d like to think the best of me is still hiding up my sleeve.”
Table of Contents

Abstract ...............................................................................................................................................ii
Acknowledgements ............................................................................................................................iv
Table of Contents ...............................................................................................................................vii
List of Tables .........................................................................................................................................ix

Chapter I: Introduction .........................................................................................................................1
  Purpose of Current Study ......................................................................................................................3
  Rationale for Current Study ...............................................................................................................3
  Research Questions ............................................................................................................................5
  Hypotheses .........................................................................................................................................6
  Definition of Terms .............................................................................................................................6

Chapter II: Literature Review ...............................................................................................................8
  Summary of Literature: Sound in Cognition & Learning .................................................................8
  Cognitive Theories Related to the Use of Instructional Media .........................................................9
    Multimedia Learning Theory ..........................................................................................................9
    Dual Coding Theory .........................................................................................................................14
    Cue Summation Theory ....................................................................................................................16
    Cognitive Load Theory ...................................................................................................................17
    Arousal Theory ...............................................................................................................................19
    Coherence and Arousal ....................................................................................................................19
  Audio Research ...............................................................................................................................20
  Music ................................................................................................................................................21
  Summary of Literature: Human Psychology and Marketing .........................................................22

Chapter III: Methodology ....................................................................................................................27
  Participants and Study Setting ..........................................................................................................27
  Principal Investigator .......................................................................................................................28
  Research Design ...............................................................................................................................29
    Study Narration ...............................................................................................................................29
  Instrumentation ...............................................................................................................................31
  Data Collection .................................................................................................................................33
  Data Analysis ....................................................................................................................................35
    Independent Variables ....................................................................................................................35
    Dependent Variable ........................................................................................................................35
Chapter IV: Findings...........................................................................................................36

Summary of Results........................................................................................................43

Chapter V: Discussion....................................................................................................44

Comparisons with Pilot Study.......................................................................................46
Limitations......................................................................................................................47
Implications for Practice...............................................................................................47
Implications for Future Research..................................................................................48

References ....................................................................................................................52

Appendix A (IRB-Approved Recruiting Announcement)..................................................60
Appendix B (IRB-Approved Consent Form)....................................................................61
Appendix C (Story Narration Script).............................................................................64
Appendix D (Retention and Transfer Test)......................................................................68
List of Tables

Table 1  Results for Retention and Transfer Tests in Audio Narration Using Different Treatments (All Questions) 37
Table 2  Results for Retention and Transfer Tests in Audio Narration Using Different Treatments (Retention Questions) 38
Table 3  Results for Retention and Transfer Tests in Audio Narration Using Different Treatments (Transfer Questions) 39
Table 4  Results for Retention and Transfer Tests in Audio Narration Using Different Treatments (Sound Effects-Related Questions) 40
Table 5  Results for Retention and Transfer Tests in Audio Narration Using Different Treatments (Non-Sound Effects-Related Questions) 41
Chapter 1: Introduction

Instructional technology is an important element of contemporary higher education. The increasing number of online and hybrid courses, along with the popularity of wireless mobile devices, have generated higher demands for recorded digital media content (Nicholas, 2008). However, many education professionals do not have a background in media production and are not familiar with the best practices and research associated with media production (Hew, 2009).

Instructors and designers can choose from pre-packaged media content that is available from textbook publishers or online sources, or they can produce their own media content. Those who choose to do the latter need to decide which mix of media to use. Video and animation can be effective tools, but they require higher-level production skills than audio (Stanley, 2006). Obviously, video and animation are visual media, and they can also contain audio and text to supplement the visual content. Because of the extra media types that are commonly included in video and animation, they also require more time in the planning and production stages. The training process for working with video and animation also requires more time and effort than learning how to work with audio (Seo, Curran, Jennings, & Collins, 2010). Audio, on the other hand, is a singular medium with no other components. For most beginners, audio is easier to learn than video and animation, and the most logical medium to begin incorporating into courses. The software and hardware investment is minimal. The user-friendliness of audio makes it a common format for instructors to distribute lecture, review, and discussion material (Hew, 2009). It is also more accessible to students because audio can be stored and played on a larger variety of playback devices (Taylor & Clark, 2010). The flexibility and portability of audio can help facilitate learning (Collis & Moonen, 2002). Visual media, which includes animation and video, can be used to complement audio (Taylor & Clark, 2010).
A research dilemma exists when learning effects of different media are compared. Clark (1983) suggested that a variety of media can be incorporated into a curriculum, and that if the instructor used effective teaching methods, each medium could facilitate learning. Subsequent media learning comparison studies contain confounding variables and conditions that make it difficult to conclude that the medium alone affects learning (Hastings & Tracey, 2005).

Institutions of higher learning have been encouraging faculty to use technology as an integral component in their instructional strategies (Bennett & Bennett, 2003), and audio podcasting is becoming a popular content distribution method in traditional classroom, hybrid, and distance learning format environments (Lum, 2006). However, educators who have not had formal training or experience in audio production can struggle with the task of assembling materials that function as effective learning tools. Unfortunately, the majority of colleges and universities do not offer formal audio production training (Seo, Curran, Jennings, & Collins, 2010). There are numerous decisions to make when producing audio, including whether or not to use music, sound effects, or both. Enhancing audio with music and sound effects requires the producer to expend extra time and effort than the amount that is required to produce a voice-only audio project. Student learning is one of the factors that faculty and instructional designers can use to decide if the extra effort of adding sound effects and music is worthwhile.

In order to gain a higher degree of knowledge about the effective use of audio, it is helpful to understand how humans process information, how different types of media mixes can affect that processing, and how a producer can use existing research to take full advantage of the power of audio to enhance learning. Listening and communication theories are useful tools for educators, but they do not fully explain human cognitive processing. Several theories and principles related to human cognition, such as Dual Coding Theory, Cue Summation Theory,
Cognitive Load Theory, Arousal Theory, and Multimedia Learning Theory, are commonly used
as the basis for audio-related studies. This paper will include a discussion of the common
theories and how they have been applied to contemporary audio research.

**Purpose of Current Study**

The purpose of the current study was to determine if the presence of different
combinations of music and sound effects in an audio-only educational narration significantly
decreased learning by measuring each participant’s retention and transfer of an educational audio
narration’s content. Existing research in multimedia learning and listener arousal and attention
will be applied to the present study. This study included two primary objectives: (a) to determine
if the sound effects and music will significantly decrease learning and (b) to determine how
different combinations of production elements compare with each other in terms of the amount
learned. Participants were assigned to one of the following based on a unique combination of
elements: a control group who heard a voice-only narration with no music or sound effects, or
one of three treatment groups who heard either a narration with sound effects, a narration with
music, or a narration with sound effects and music.

**Rationale for Current Study**

Instructors and instructional designers can produce content using a variety of media,
including video, audio, and animation. In order for the media content to be an effective learning
tool, instructors and instructional designers should possess introductory production skills. A
substantial body of research has been compiled on the use of sound effects and music in
animated educational narrations and in radio commercials, but extensive research has not
measured them in audio-only educational narrations. Much of the existing research on
audio-only educational narrations measured either student satisfaction or test scores without
considering sound effects or music. Findings from this study will begin to bridge the current research gaps by applying concepts from existing research in animations and radio commercials. Education professionals, who include teaching faculty and instructional designers, can benefit by knowing whether or not to put in the extra time that is necessary to include sound effects and music. By using this study’s findings, time can be saved and more effective audio can be produced.

The present study drew in part from the Coherence Principle of Multimedia Learning Theory (Mayer, 2009), which measures the effects of extraneous material on learning in multimedia presentations. This study also incorporated information from arousal theory experiments, which measure different ways to attract listeners’ attentions through the use of audio production elements. Cognitive and psychological theories form the basis of the Coherence Principle and arousal theory studies, but, as previously stated, they have limited their analysis to animations and radio commercials, respectively.

The Coherence Principle and Arousal Theory are conflicting paradigms. The former posits that sound effects and music hinder the learning process (Mayer, 2009), while arousal theory contends that external stimuli keep listeners’ attention and help them remember information (Kintsch, 1980). The evidence of improved information retention in the radio commercial studies suggests that more research is needed in the education disciplines. There might be some differences in Mayer’s findings with animations when compared with audio-only narrations.

Additionally, the current study hopes to clarify some of the discrepancies found within existing theoretical research. According to Barron and Callandra (2003), applications and interpretations of media-related learning theories differ and their results are contradictory. While
some studies have shown evidence of modality effects, where there are significant differences among treatments of audio elements (Kalyuga, Chandler, & Sweller, 1999; Mayer & Moreno, 1998; Moreno & Mayer, 1999; Tindall-Ford, Chandler, & Sweller, 1997), others have shown no modality effects (Barron & Atkins, 1994; Barron & Kysilka, 1993; Curran, 2011; Koroughlian & Klein, 2000). Such discrepancies point out the need for continuing research to help identify specific conditions that can be tested in media research.

Finally, applying Mayer’s Coherence Principle 2 research to audio-only educational non-scientific content with adult participants is a viable worthwhile project because of the opportunities that remain for research in this area. Mayer (2009) acknowledged that his area of interest and expertise is scientific content, and he called for research in other topics. Hew (2009) found that instructors in the English discipline use podcasts less frequently than instructors in other disciplines, possibly due to lack of knowledge about the production aspects. Mayer (2009) also noted that his research has been conducted with K-12 learners, and encouraged researching MLT principles with adult learners.

**Research Questions**

Three research questions guided this dissertation study:

1) How does the addition of sound effects affect learning when compared to a voice-only narration?

2) How does the addition of background music affect learning when compared to a voice-only narration?

3) How does the addition of a combination of sound effects and background music affect learning when compared to a voice-only narration?
Hypotheses

The current study replicated three research hypotheses related to Coherence Principle 2, but in an audio-only narration, rather than in a narrated animation. The research hypotheses were:

H₁: Learning is improved when sound effects are excluded from an audio-only educational narration.

H₂: Learning is improved when music is excluded from an audio-only educational narration.

H₃: Learning is improved when music and sound effects are excluded from an audio-only educational narration.

The null hypothesis was:

H₀: Learning is not improved when music and/or sound effects are excluded in an audio-only educational narration.

Definition of Terms

The following terms are defined for the benefit of readers who are not familiar with them.

Arousal. This commonly refers to the activation of the nervous system in response to audio elements such as voice, sound effects, and music.

Cognitive Load Theory. This is a cognitive theory proposed by John Sweller in 1979. According to this theory, the amount of human working memory is limited, so each learner must use working memory efficiently.

Coherence Principle. Richard E. Mayer proposed this three-part principle that is a component of Multimedia Learning Theory (see below). The Coherence Principle states that people learn better when extraneous material is excluded rather than included. Part 2 of the
Coherence Principle, which is relevant to the current study, states that learning is improved when interesting but irrelevant sounds and music are excluded from a multimedia presentation.

*Cue summation.* Hartman proposed this cognitive theory in 1961. According to this theory, learning increases as the number of available cues or stimuli is increased.

*Dual coding.* Allen Paivio proposed this cognitive theory in 1971. According to this theory, both visual and verbal information is used to represent information, and that these two distinct types of information are processed differently.

*Multimedia Learning Theory.* Richard E. Mayer proposed this cognitive theory in 2001, and updated it in 2009. This theory is centered around 12 principles that address the effectiveness of different multimedia presentation strategies on learning. This theory is defined as being learner-centric because it examines how humans process information.

*Narration.* Narration is a written or spoken account or story, which is often a recital of chronological events.

*Retention.* Retention is the act or power of remembering things.

*Sound effect.* Except for music or speech, this refers to any sound in an audio presentation. It can be recorded naturally in its working environment or artificially reproduced to create an effect in a dramatic presentation, such as the sound of a storm or a creaking door.

*Transfer:* A person transfers information by demonstrating the ability to apply learned content to problem solving and critical thinking activities.
Chapter II: Literature Review

The current study drew from different theoretical frameworks of audio research, including the effective use of audio education and in business. A review of studies in both of these general areas is necessary in order to fully understand the context of the current study. The first part of this review of literature presents information about the impact of sound on learning and cognition. Next, the literature about the underlying theories guiding this dissertation is presented. These include Multimedia Learning Theory, Dual Coding Theory, Cue Summation Theory, Cognitive Load Theory, and Arousal Theory. The next part of this chapter discusses literature related specifically to audio research, followed by a discussion of media research in the fields of psychology and marketing. The chapter concludes with a brief summary and analysis of the existing literature.

Summary of Literature: Sound in Cognition & Learning

There have been various educational studies in audio that measured learning by modality effects, in which different media combinations were compared in attempts to determine if some were more effective than others. Most of the studies evaluated the effectiveness of audio in multimedia presentations, with only a small number evaluating audio as the sole medium in a presentation. In the latter category, the results for audio effectiveness were mixed. Kalyuga, Chandler, and Sweller (1997) tested an industrial training module using different media combinations, concluding that the audio-only version was superior to text-only and an audio-text combination. The superiority of audio over text has been supported (Clark & Mayer, 2003; Mayer, 2009; Mayer & Moreno, 1998; Moreno & Mayer, 1999). With slightly different findings, Mayer (2009) also concluded that words and pictures are more effective than words alone.
The effectiveness of audio over text also yielded inconclusive results in experiments using multimedia presentations. For instance, Tindall-Ford, Chandler, and Sweller (1997) conducted three audio-visual experiments with adults, and they concluded that listeners can integrate the difficult content easier if it is an audio format versus a written format. They also concluded that there was no significant difference in learning with audio versus text when lower-level content was presented. In the previously cited audio studies, the researchers controlled the pace of learning. Barron and Calandra (2003) analyzed seven studies that compared different media combinations in multimedia presentations. After aggregating the results of the studies, they concluded that neither the presence nor absence of audio had a significant effect on learning. Tabbers (2002) conducted a series of multimedia presentation experiments that allowed the user to control the pace of learning. In those studies, text was found to be superior to audio.

Cognitive Theories Related to the Use of Instructional Media

Instructional media theories began to surface in academic research as the use of instructional media gained popularity. They formed the basis for much of the contemporary empirical studies that measure learning in different testing conditions. The theories that are relevant to the current dissertation study are those that examined how the human senses receive and process information. In particular, the current study was rooted in theories that examined information overload, distractions, arousal, and channel capacity.

Multimedia Learning Theory. In a well-cited theory for media development and learning, the Multimedia Learning Theory (MLT) (Mayer, 2009) is an important framework for examining research on audio in educational presentations. Several other theories related to learning though multiple channels and modes were present prior to Multimedia Learning Theory (Barron & Calandra, 2003), which incorporates some of their elements. MLT was derived from
earlier cognitive theories such as Cue Summation (Hartman, 1961), Dual Coding (Paivio, 1971), and Cognitive Load (Sweller, 1988). MLT also cited arousal theories to illustrate the contrast in their general premises. In turn, MLT has spawned articles and studies that examine the use of audio with other media such as text and animation (Barron & Calandra, 2003). Although this theory concentrates on the use of audio when combined with other media, it also provides a base of knowledge that can be applied to the study of audio-only presentations. Prior to the introduction of MLT in 2001, multimedia presentation designs were often centered on technical aspects of media, rather than instructional research or theories (Barron & Atkins, 1994; Moreno & Mayer, 1999).

Overall, MLT examines learning outcomes of presentations using different combinations of written and spoken words, and pictures (Mayer, 2009). The theory, which is rooted in human cognitive science, evaluates the use of multimedia via learner-centric views. Instead of monitoring the listener’s behavior, MLT measures the cognitive activity during the learning activity, and incorporates concepts from dual coding and cognitive load studies. According to Mayer (2009), the study of cognitive activity is more useful because it measures meaningful learning, while behavior observation only measures rote learning. The general premise for MLT is that the brain can only effectively process so much information at one time. MLT is based on theories such as Dual Coding (Paivio, 1986), who indicate that humans have a dual-channel system for processing information. In dual-channel processing, visual and aural information compete for cognitive memory, and they must be presented in such a manner so that learners can effectively process the content into meaningful learning. Mayer (2009) claimed that instructional designers who focus on multimedia technology without considering the human sensory system
will not foster meaningful learning as well as those who understand how humans process information.

As a theory, MLT has twelve major principles that were fully tested through a series of experimental designs (see Mayer, 2009). The Coherence Principle is one of five principles that are concerned with reducing extraneous processing, and it is the only one of the 12 that studies the inclusion of sounds and music in a multimedia presentation. The principle is divided into three complementary parts, all of which hypothesize that excluding extraneous material from a multimedia presentation improves learning. Mayer (2009) identifies the parts as Coherence Principle 1, Coherence Principle 2, and Coherence Principle 3. Coherence Principle 1 hypothesizes that interesting but irrelevant words and pictures can significantly reduce learning. Coherence Principle 2 hypothesizes that interesting but irrelevant sounds and music can significantly reduce learning. Coherence Principle 3 hypothesizes that unneeded words and symbols can significantly reduce learning. The theoretical rationale for the Coherence Principle is based on the premise that extraneous material competes for cognitive resources in working memory, which might divert the learner’s attention to important material. In turn, the extraneous material could disrupt the learner’s ability to organize the content (Mayer, 2009).

According to Mayer (2009), multimedia learning assumes that humans process information through a dual channel system: one for processing visual information and one for processing auditory information. Each channel has limited capacity, and active learning occurs when the cognitive processes inherent in both channels are coordinated. As previously discussed, Cue Summation Theory (Hartman, 1961) states that more learning occurs in presentations that engage dual channels than those that engage a single channel due to the increased number of stimuli that the learner receives, and Dual Coding Theory (Paivio, 1986) states that information
is processed either verbally or non-verbally. Related studies such as Hannafin and Hooper (1993) concluded that dual coding is not effective when multiple sources of information engage the same channel. For example, sound and text are both processed verbally, so presenting identical words simultaneously in both narration and text form would be inefficient and should be avoided. In another widely cited theory, the Cognitive Load Theory (Sweller, 1988) states that humans have limited working memory and capacity to retain information. Proponents of Cognitive Load Theory advise designers to minimize the stress on the typical learner’s short-term memory by using multiple modalities, reducing the complexity of the material, and organizing the material in a logical manner.

From a learning context, the student-centered approach of Multimedia Learning Theory is based on how the human mind works rather than the technology that is available. The principles that comprise Multimedia Learning Theory advocate using multimedia technology as an aid to human cognition. If a multimedia design considers how the human mind works, it will be more effective in its ability to promote learning than designs that do not take the human mind into consideration (Mayer, 2009).

For instance, experiments using animations with non-adult participants supported Coherence Principle 2. Moreno and Mayer (2000) conducted two studies to measure the effect of sounds in two narrated animations. One animation explained how lightning is formed, and the other described how a car’s braking system worked. Each study divided the participants into two groups. The control groups viewed their respective animation absent of sound effects and music. The treatment groups viewed their respective animation with sound effects and music. In both experiments, the control groups’ transfer test scores were significantly higher than those of the treatment groups.
**Coherence Principle 2 limitations.** In a discussion of the Coherence Principle 2 boundary conditions, Mayer (2009) acknowledged that more experimentation is necessary, especially with adults. The boundary conditions discussion also suggested that sounds might be appropriate for narrations that contain emotional content that stimulates positive or negative attitudes, and for narrations where the music and sound effects are part of a topic’s essential, lower-level content. Emotional stimuli could also impair or facilitate memory (Kensinger & Corkin, 2003; Perlstein, Elbert, & Stenger, 2002). Emotional messages that stimulate arousal can increase memory (Bolls, Lang, & Potter, 2001). Sounds also tend to be less of a distraction for higher-level and high-memory capacity learners (Mayer, 2009).

Coherence Principle 2 studies have other limitations that could lead to different conclusions about using sound effects and music. In both experiments, Mayer (2009) worked with science-based content. Participants could have had existing knowledge of the material which would have distorted the results. However, it was not clear if participants were pre-screened to determine their background knowledge. The research discussion did not provide details about how the sound effects and music were implemented in terms of frequency, type, and variety. Other than stating that the participants did not report problems with the volume, there is no way to discern from the discussion if the sounds were produced and added to the narration effectively.

All of the Coherence Principle 2 studies used animations, and did not consider audio-only narrations. In an audio-only narration, the visual channel will not be taxing listeners’ cognitive resources, so more attention can be devoted to the narration. The auditory channel has more capacity than the visual channel, and auditory information is processed more efficiently in a single-mode presentation than in a dual-mode presentation such as an animation (Ronnberg &
Ohlsson, 1980). For instance, Bolls & Muehling (2007) confirmed that audio messages are not processed as effectively when they are competing for cognitive resources with the visual channel.

It can be difficult to fully evaluate empirical research when independent variables are not isolated. Another limitation of Coherence Principle 2 is that the narrations for the treatment groups in the studies contained sound effects and music. The music was present throughout both treatment groups’ narrations, while the sound effects were mixed over the music at different intervals. The sound effects always played simultaneously with music. Neither element was isolated in the animations, so it would be difficult to postulate whether or not one of those elements was responsible for the decreased learning. For example, Moreno and Mayer (2000) conducted an experiment using a multimedia lesson with four different treatments of audio: narration only, narration with sound effects, narration with background music, and narration with sound effects and background music. The results showed that the narration-sound effects-music group recalled a significantly less amount of information than the other groups, the narration-music group recalled a significantly less amount of information than the narration and narration-sound effects groups, and that there was no significant difference between the narration and narration-sound effects group. Again, sound effects and music were not addressed separately in Coherence Principle 2. The principle might have been presented differently if the experiment design had tested sound effects and music separately. Educators who only read Coherence Principle 2 but did not read the study by Moreno and Mayer (2000) would not be aware that sound effects might have more value than music in multimedia presentations. Research is needed to study the effect of multiple simultaneous enhancements on cognitive processing in the audio channel.
**Dual Coding Theory.** When multimedia presentations are produced, the creator can increase listener engagement by understanding how people process information. The human cognitive system processes information through two channels: the visual and the auditory. When a person watches a multimedia presentation, the visual channel interprets video, animation, and graphic elements such as illustrations and photographs. The auditory channel processes audio and text elements. Although text appears on the screen, the audience reads the information and speaks the words aloud or to themselves as they are read. Each channel has a limited processing capacity.

Paivio (1971) proposed that people process verbal and non-verbal information separately and with equal weight, and that verbal knowledge representation is qualitatively different from visual knowledge representation. This process of *dual coding* is more efficient if each source of information is coded in separate channels. The two verbal components, voice and text, should not be presented simultaneously because they can put too much overload on the verbal channel (Hannafin & Hooper, 1993). Visual components such as video and photographs can be presented simultaneously with one of the verbal components (Barron & Calandra, 2003). Thus, the amount of information going into a channel must be regulated. The same concept holds true for the number of simultaneous pieces of information. If too many different elements occur in the same channel, the audience will not be able to process all of it effectively.

Designing research examining how to balance input, which is based on the dual coding theory, examines the amount of audio and visual objects in a presentation (Sadoski, Paivio, & Goetz, 1991). If the researcher believes that there are too many visual elements, some of that content would be deleted or transferred to the audio portion (Mayer & Moreno, 2003). Similarly, too much audio content would overload the auditory channel.
In an audio-only presentation, the absence of video will limit the amount of information being processed in the visual channel. Likewise, the absence of text will free up more capacity in the auditory channel so that the listener can process the information more efficiently. However, the channel can still become overloaded, which will be determined by a number of factors. One of the factors is how many different types of audible elements are added to an audio piece, the intervals between each new element, the placement within the production, the volume, and the purpose of the element.

**Cue Summation Theory.** The Cue Summation Theory predicts that learning increases as the number of stimuli increases (Hartman, 1961). For example, an audio presentation with a simultaneous print presentation is more effective than either of those media presented alone. Cue summation research assumes that the information being presented is redundant, so that the printed content is the same as the simultaneous audio content. However, earlier research theorized that multiple communication channels are not advantageous because only a single channel links senses to the central nervous system (Broadbent, 1958).

The body of research on cue summation has been inconclusive. Van Mondfrands and Travers (1964) concluded that no significant differences in learning existed between single and multiple channels, concluding that previous studies suffered from small sample sizes, lack of controls, and the lack of significant tests. Severin (1967) concurred that multiple channel research has produced mixed and sometimes contradictory research because of issues such as poor sampling, weak design, and invalid or unreliable tests. Other reviews suggested that human characteristics were more likely factors in learning differences. Allen and Cooney (1964) suggested that age and maturity are better predictors of learning than the number of channels is a
presentation. Hsia (1969) studied modality and intelligence levels, and concluded that lower-level learners benefitted from a limited number of inputs.

Cue summation and single-channel theories are both strengthened when they are placed in context. In an attempt to reconcile Hartman’s and Broadbent’s conflicting theories, Hsia (1968) proposed that learning through multiple channels was more effective as long as the limits of the central nervous system were not taxed. Extraneous material would be more suited to a single-channel presentation.

Consistency across cue summation studies is another issue for researchers. Reinwein (2011) cautioned about comparing the results of studies using text and visual elements to studies using spoken words and visual elements. The human listening and reading processes are different, both in the learning methods and in practice. However, Reinwein (2011) suggested that comparing the results of text and visuals to words and visuals could be useful if the same illustrated text is used. There has been a lack of research using that approach.

**Cognitive Load Theory.** Instructors and producers who produce audio need to be cognizant of the length and complexity of each content unit. Each person’s cognitive system has a threshold that will be exceeded when a task becomes unmanageable. The limit, which has been identified as cognitive load, will be reached when the task puts too much of a burden on the cognitive system (Sweller, 1994). Every task will exact some toll. The normal burden, known as intrinsic load, is inherent to every task (Verhoeven, Schnotz, & Paas, 2009). Intrinsic load is determined by the person’s expertise and motivation with no instruction (Verhoeven, Schnotz, & Paas, 2009). Learners with more background in the specific area of expertise will be able to store a higher level of intrinsic load, which means they will be able to complete more of a task without instruction. For example, a seasoned mechanic will have no problem changing the oil in
a car, but a novice do-it-yourselfer might struggle with the job. A similar scenario occurs in the classroom. When instruction is introduced to non-expert learners, extraneous cognitive load will increase, which would not be true of expert learners. Well-designed instruction will minimize extraneous load, but poorly-designed instruction will increase it to a level that requires irrelevant and inefficient cognitive processing (Sweller & Chandler, 1994; Tempelman-Kluit, 2006). Well-designed instruction related to cognitive load can also be referred to as germane instruction (Verhoeven, Schnitz, & Paas, 2009), which is the instruction that helps learners progress from novice to expert.

The traditional view of cognitive load research is to explore methods of reducing extraneous load (Mayer & Moreno, 2003). Researchers have not given similar attention to studying ways to increase germane load, which is the cognitive load that results from self-learning and memorization of material as the learner progresses from novice to expert. When intrinsic load is low, the learner should be able to understand introductory subject matter. Thus, there should be enough unused memory that can accommodate germane instruction (Schnitz & Kurschner, 2007). Suggestions have been proposed that encourage instructors to minimize extraneous load and maximize germane load, but Schnitz and Kurschner (2007) suggested that germane load be adjusted to fit the learning task. The instructor would be responsible for determining the amount of instruction that each student will need to achieve the maximum germane load.

Working memory affects the storage and processing of information. Each person has a different working memory capacity, so it is difficult to quantify a generic maximum level. Web-based learning environments require students to possess certain digital skills. Researchers have
questions and concerns related to the cognitive load these skills will impose on students (van Gog & Paas, 2008).

Designing for cognitive load has a goal of providing the proper amount of material for students (Mann, 2008). Too much material is known as overloading; too little material is known as underloading. Instructional designers and instructors who combine visual elements with complementary audio elements can create cognitive overload and could possibly interfere with learning (Mann, 2008). Researchers who designed for audio-visual sensations found that sound needs to be more prevalent if students are being inundated with many visual sensations or they might not be able to process the information (Burkman, 1987). Instructional designers who designed for structured sound function focused on making sure that the sounds that accompany visual elements have a purpose (Mann, 2008).

Cognitive load and its related design strategies are important considerations for instructors and designers because if they are not factored into the design process, other design factors of the presentation might be negated.

**Arousal Theory.** Arousal Theory is related to human emotions and motivation, and it has been applied to educational studies in different disciplines, including Instructional Technology Design. It runs contrary to MLT because it posits that students learn better when they are emotionally aroused by interesting material such as music and sound effects.

**Coherence and arousal.** Sound effects and music are often used to enhance audio messages. There are conflicting opinions on their value. The assumption is that the enhancements capture listeners’ attention and help them retain the message more effectively. However, humans’ auditory channels have limited capacity, which restricts the amount of information that can be processed (Moreno and Mayer, 2000). Coherence theory supports the
idea that retention and transfer increase when fewer auditory elements are used in an audio presentation (Mayer, 2009). On the contrary, arousal theory supports the idea that such elements increase retention and transfer because they make the presentation more interesting, arousing the listener’s senses and increasing their attention (Mayer, 2009). Moreno and Mayer (2000) examined arousal and coherence theories in their study.

Audio Research

Coherence Principle 2 of Multimedia Learning Theory focuses on animations, which contain a visual component. MLT, however, does not examine audio-only presentations. Studies have been conducted on audio-only media that have begun point out how people use these presentations differently from animated media. These studies have taken place outside of education in areas such as advertising and radio broadcasting. In education, however, very little research has paid attention to the effects of audio on learning and attention. When examining the educational audio research literature, an obvious research gap emerges.

Hew (2009) summarized a body of research encompassing 30 studies on instructional audio-only podcasts. The studies were mainly in the engineering, science and technology, and business and law disciplines. The summary indicated that the studies focused on either student satisfaction or test scores, but none examined the use or effect of production elements. The studies commonly compared groups who listened to the podcasts and those who did not listen. One of the findings of this review is that students enjoy listening to instructor-produced podcasts, indicating a need for them, particularly in online and hybrid courses.

Another body of research introduced extensive analysis of the psychological aspects of media effects. This area of study has become popular because it probes the mental processes of receiving messages from the media specifically focused on marketing (Lang, Bradley, Chung, &
Lee, 2003). The studies used professionally produced radio commercials with different mixes of music, sound effects and voices to gauge listeners’ reactions to the production elements, their recall of information in the commercials, and their levels of arousal after hearing specific production elements. The consensus of these studies was that production elements can stimulate arousal and attention, which can help improve memory. For example, in an extensive study using nine different audio elements, Potter, Lang, and Bolls (2008) found that listeners retained information about a commercial after a different production element was introduced, and also that it could take a few seconds for listeners to become fully oriented to a new production element. Arousal can help listeners’ increase their recall memory (Potter & Choi, 2006).

The radio commercial studies can help bridge the research gaps in educational audio research. The same experimental designs that are used in those studies can be replicated in studies using educational content. By going beyond student satisfaction surveys and perceived student learning questionnaires, researchers will be better able to understand effective production methods for educational audio.

**Music.** Arousal theory assumes that extraneous material such as music will arouse the senses and help learners pay more attention to the primary material, which is narration (Mayer, 2009). Several studies were conducted to compare the effect music had on student engagement. Although the majority of cited studies incorporated visual presentations of learning content, and some required participants to memorize sequences, they present useful research as a prelude to the current study, even though the results were inconsistent.

In a widely-cited study, Salame and Baddeley (1989) implemented three methods that measured the effects of instrumental and non-instrumental music and sound effects on the ability to recall visually-presented number sequences. The first two methods compared music to a silent
condition. The third trial added sound effects as a condition. In the first and third trials, both styles of music significantly decreased recall. In the second trial, which used different music, instrumental music did not significantly decrease recall. However, vocal music did significantly reduce recall. Iwanaga and Ito (2002) used a similar recall design, substituting word patterns and artistic patterns for number sequences. There was no significant difference among the music and non-music groups in the verbal task of recalling word patterns, but music did significantly reduce recall in the spatial task of recalling spatial patterns. Alley and Greene (2008) tested two groups, each of whom listened to all treatments of no sound, sound effects, instrumental, and vocal music. The groups’ means were not significantly different for the same treatment. However, when the groups’ scores were aggregated by treatment, music significantly decreased recall. In all of the previously cited music studies, instrumental music recall scores were higher than scores in the vocal music group. In one case, the difference between music groups was significant (Alley & Greene, 2008). In another case, the difference was not significant (Iwanaga & Ito, 2002). In a multiple method experiment there were mixed results (Salame & Baddeley, 1989).

The inconsistent results in the previously sighted reports used participants from a random sample with no specific qualifying characteristics other than being students. A stratified sample, such as students with a specific learning disorder or medical condition, might yield different results. In a study employing a small group of students with autism, Carnahan, Basham, and Musti-Rao (2009) found that adding music to interactive books improved engagement for students with autism during small group instruction.

Summary of Literature: Human Psychology and Marketing

Looking at the issue from a different perspective, the psychological approach to studying media effects has become popular because it probes the mental processes of receiving messages
from the media specifically focused on marketing (Lang, Bradley, Chung, & Lee, 2003). The medium of radio has provided opportunities for studying how production elements are cognitively processed, which can affect the listening experience. Commercials routinely contain mixes of voice, sound effects, and music designed to engage the listener’s imagination (Bolls & Lang, 2003). Part of the imagination process involves imagery, which happens when listeners use one of their other senses to experience what is being conveyed in the commercial message (Bolls, 2002; Kosslyn, 1994; Loverock & Modigliani, 1995; MacInnis & Price, 1987). For example, a commercial for a steak house can cause a listener to figuratively “smell” the smoke of a searing sirloin on a grill. Sound effects in a commercial promoting a concert can imaginarily “transport” someone so that the person virtually experiences the thrill of watching the event from the first row. A descriptive commercial for a golf course can stimulate the sense of touch to the point that the listener imagines “gripping” and “swinging” a club while the commercial is playing.

Professionally produced radio commercials have some level of auditory structural complexity, which is quantified by the number well-planned production elements that can cause listeners to react, such as rapid voice changes, sound effects, and music (Potter & Choi, 2006). Researching listener reaction has been well studied. For instance, research on children’s television programs observed viewers’ reactions to the introduction of unexpected sounds during different programs. In their work, Watt and Welch (1983) found a positive correlation between high audio complexity and children’s ability to remember educational TV program content, which was different from Mayer’s findings. They believed that the orienting response (OR) built into humans was responsible for the relationship because the OR attracts the attention to signals
such as audio. Orienting responses are often followed by message encoding (Geiger & Reeves, 1993).

As no to little surprise, some radio commercials are produced with the intent of eliciting images in each listener’s mind (Bolls & Muehling, 2007). In fact, commercials can also be classified by level of imagery. Radio commercials that contain high levels of imagery can activate both the image and verbal processing channels (Bolls, 2002; Goosens, 1994). The verbal system semantically processes the words while the image system converts these words into mental images (Bolls & Lang, 2003). Animations already have a built-in visual element that occupies much of the visual channel’s capacity, leaving precious little room for processing extraneous sounds (Bolls & Muehling, 2007). However, it is thought that audio-only recordings do not have a natural visual element, so the visual channel has more available capacity to process images whenever a sound effect activates that channel. With this line of thought, listeners should then be better able to process sounds with audio-only recordings than they would by watching an animation.

In fact, there is evidence that sound effects can stimulate arousal and attention in radio commercials (Bolls & Lang, 2003; Bolls, Lang, & Potter, 2001; Bolls & Muehling, 2007; Potter & Choi, 2006). Studies from television research have been used as a basis for similar radio research, Potter and Choi (2006) also found that high levels of arousal positively correlate with free recall of message content in radio commercials. Television commercials that contain more changes in production elements, such as rapid video edits, tend to cause more arousal. This increased arousal is likely to trigger more processing resources that are required to store messages in long-term memory. Emotion can also aid memory because radio commercials that impart highly emotional content tend to increase arousal (Potter & Choi, 2006).
In radio commercials, listeners reported a more positive attitude to the non-claim portions of high-imagery recordings than those categorized as low-imagery (Potter & Choi, 2006). Non-claim portions do not contain messages that include an advertiser’s claims about their product, service, or business. There is an indirect connection between attitude and memory of non-claim content in radio commercials. High-imagery commercials tend to foster favorable attitudes and increased memory. Potter and Choi (2006) did not measure how attitude might affect memory, but Brosius (1993) found that viewers tended to remember more information about TV news stories when they had a positive attitude about those stories. Mayer (2009) did not categorize audio by level of imagery, so no inferences can be drawn about attitude from the Coherence Principle studies. Attitude was not included in any of the Multimedia Learning Theory Principles’ hypotheses.

There seems to be some disconnect between the two main approaches to audio research. The methods are not consistent. Much of the literature in the cognition and learning arena measured test scores or participants’ perceived learning levels. Some of the studies measured the effects of music and sound effects in broad terms, rather than testing multiple attributes such as music tempos or genre. Some of the studies in human psychology and marketing were quite detailed about describing the different types of music and sound effects that were tested. Some of the psychology and marketing studies measured memory, but the focus on some of these studies was simply arousal. Some of the studies in cognition and learning and human psychology and marketing acknowledged that more detailed multivariate methods should be implemented in future research. The researchers in each area of study could borrow methods from each other or even collaborate on future projects to help strengthen the breadth of knowledge in audio research.
The results between some of the studies were inconsistent, and in some cases, contradictory. In many of the Multimedia Learning Theory-related studies, the results indicated that adding sound effects and music decreased learning. However, many of these studies used animations, which contain a visual component. The extra memory required to process visual content could have been the reason why the music and sound effects decreased learning. In the audio-only psychology and marketing studies, the results indicated that extra auditory elements aroused listeners and helped them to better remember information that was presented. Many of these studies used non-educational content and were shorter in length, so it was not apparent that the results would be similar if a study tested longer educational audio content. The purpose of this study was to clarify some of the discrepancies in the different types of studies by applying the Coherence Principle 2 studies methods to an instructional audio study.
Chapter III: Methodology

Participants and Study Setting

This study was a quantitative experimental design that employed a survey instrument and had four conditions. The purpose of the study was to determine if the presence of different combinations of music and sound effects in an audio-only educational narration significantly increased or decreased learning by measuring participants’ retention and transfer test scores.

Three research questions guided this dissertation study:

1) How does the addition of sound effects affect learning when compared to a voice-only narration?

2) How does the addition of background music affect learning when compared to a voice-only narration?

3) How does the addition of a combination of sound effects and background music affect learning when compared to a voice-only narration?

The study setting was in one of the five standard computer laboratory classrooms that the university designed on the campus. It houses 24 student computers contained in a four-row configuration, with six computers per row. All computer workstations are Windows-based, each with 17-inch flat screen monitors, a mouse, a keyboard, and built-in speakers, which were muted for this study because the participants used headphones. The room was designed to be well-lit and spacious. The instructor workstation is at the front left side of the room and is connected to a projector screen. The study setting was in a quiet area in a side hallway that does not get a heavy amount of foot traffic. Thus, the outside noise did not encroach on the study setting.

The participants were college students ($n = 143$) enrolled in seven sections of five different Liberal Arts courses at a large Midwestern public university. The gender breakdown
was 84 females, 58 males, and 1 unreported. The age breakdown was 86 students in the traditional college student age range of 18-24, 56 aged 25 and over, and 1 unreported. A power analysis suggested a total $n = 112$, which is a value of $n = 28$ for each group. The Principal Investigator set a target of 120-160 participants to account for outliers, bad data, and also to strengthen the sample. The Principal Investigator sent e-mails to full-time faculty at the university to solicit interest in the experiment. The Principal Investigator provided a written and audio version of a recruiting announcement to interested instructors, who posted them on their course section websites. The script of this announcement is included as Appendix A. The instructors gave each participant extra credit in their respective course section. Of the 143 participants, 36 were randomly assigned to the control group (voice-only narrations), and the rest were assigned to one of the three treatment groups: narration with sound effects ($n = 36$), narration with music ($n = 36$), and narration with music and sound effects ($n = 35$). Participants were not allowed to take part in more than one trial.

**Principal Investigator**

The Principal Investigator was both a doctoral student in the Instructional Design and Technology doctoral program and an Associate Professor in Interactive Multimedia Technology. Prior to working at the university, he spent 25 years in the radio business, producing radio commercials for advertisers in medium-sized markets in the Midwestern and Southeastern United States.
Research Design

The current study replicated and expanded on a pilot study (Curran, 2011) that measured learning using a control group (voice-only narration) and one treatment group (sound effects). The Principal Investigator narrated and recorded a six-minute public domain short story by an anonymous author about a woman who encounters a suspicious-looking man on a public transportation bus.

Study narration. The Principal Investigator found the story in an online archive of unpublished short stories. This story was selected for a number of reasons. It had not been published in book form, so it was unlikely that any participants would have had prior knowledge of its content. The Principal Investigator asked three English Literature experts who were members of the English Department at the university to confirm the story’s appropriateness for a college student. These experts believed the storyline to be clear and easy for a college student to follow. A group of seven non-participant college students at the university listened to the story prior to the scheduled trials in order to gauge appropriateness and clarity of the passage. All of them confirmed the English Department faculty’s assessment that the story was appropriate for college students. The length also fit with the Principal Investigator’s time limit of under 10 minutes. The story, which was told from a third-person narrator’s perspective, was six minutes. The narrator read it at a normal reading pace. There were natural pauses during the narration. A master copy of the story script, which includes notations for the instances of music and sound effects, is included as Appendix C.

The Principal Investigator produced four digital audio copies of the narration in the WAV format, which is a common universal audio file type. The first version was voice-only with no enhancements. The second version was a copy of the voice-only narration enhanced with 16
different sound effects. The length of the sound effects ranged from 1 sec to 5 sec. Each sound effect was inserted at a point in the narration where the content matched it. For example, a sound effect of a microwave oven bell played just after the narrator said, “She kept it simple: canned soup and microwave Chicken Alfredo.” The third version contained voice and music. The music was a medium-tempo smooth jazz instrumental that played throughout the narration. The music did not contain any noticeable tempo or volume changes. This type of music tends to be relaxing, which is likely to reduce music tension and allows the listener to focus more on non-musical elements (Bartlett, 1996). The fourth version contained a combination of the same music and sound effects, played at the same volume as in the other versions. In all cases, the sound effects and music were played at a volume that was audible, but not distracting. A group of 10 non-participants sampled each version and reported that the volume on each narration was satisfactory. One copy of each version was deployed to computers in alternating rows at the testing site. The room contained 24 computers, so six copies of each version were available.

The six-minute narration length was chosen because research indicates that students’ listening spans for podcasts often last from 5 to 20 minutes, based on observations (Muppala & Kong, 2007). Other researchers canvassed students via questionnaires and found their preferences to be in the 3-to-10-minute range (Chan & Lee, 2005). The lower end of the ranges was chosen to lessen the likelihood of listener fatigue and dissatisfaction while allowing for a substantive amount of content to be included.

The current study used the English literature discipline for the content partly because Mayer (2009) suggested non-science disciplines for future research. The content of a fictional short story is unique, so if none of the participants was familiar with it, pre-existing knowledge cannot skew the retention and transfer test results. Short stories are more likely to feature
emotional content than fact-based material. Mayer (2009) suggested that sound might work better with emotional content, and that this area needs to researched.

The current study was not primarily measuring differences in gender or age, nor was it measuring listener fatigue. However, the Principal Investigator ran post-hoc tests to measure gender and age differences to determine if those demographic-related independent variables needed to be tested in future audio studies. The Principal Investigator also attempted to detect listening fatigue during the trials by observing participants’ facial expressions and body language.

**Instrumentation**

The Principal Investigator prepared a 20-question multiple choice test, which included 12 retention questions and 8 transfer questions related to the story. Each question contained four choices, and each question’s set of choices contained only one correct answer (see Appendix D). The 20 graded questions were presented in the same order that the related content was introduced in the narration. The Principal Investigator used this strategy because he thought it would be more consistent than asking questions out of chronological order. Specifically, the Principal Investigator believed that it would be too easy for the participants to remember information that was presented at the end of the story if questions about that content were asked at the beginning of the test. Likewise, the Principal Investigator believed that it would be too difficult for the participants to remember information that was presented at the beginning of the story if questions about that content were asked at the end of the test.

The final version of the test was pre-screened for validity and reliability. The Principal Investigator asked three members of the English department faculty at the university to review the questions to determine if the questions were relevant to the story’s content and if the
questions were appropriate for college-level English courses. All of the English faculty who assisted in the review validated the test by affirming both attributes. In a simulation of actual trial conditions, a sample of seven non-participant college students listened to the voice-only version of the narration and immediately took the test. They provided information about test question clarity. Inter-rater reliability was addressed through the use of multiple-choice answers, so there would be no ambiguity in attending consistent scores. The scores of the current study \((M = 18.22)\) were consistent with the trial results.

The test sheets for the control group were coded with the letter ‘V’ in the upper right-hand corner. The test sheets for the treatment groups had the following code symbols: “VS” for narration and sound effects; “VM” for narration and music; and “VMS” for the combination of sound effects and music. The Principal Investigator did not explain the code on the control group test sheets to the participants. Some questions related to content that was narrated immediately after sound effects were played to help determine if the sound effects caused a distraction and hindered learning.

The multiple choice format was selected over the short answer format for two reasons: to expedite the grading, data tabulating, and data reporting processes; and to increase the reliability of the instrument by eliminating the chance of the Principal Investigator misinterpreting answers during the grading process. The multiple choice format was selected over the true/false format to lessen the effects of guessing. A master copy of the test is included as Appendix D. The 12 retention questions were numbers 2, 4, 5, 7, 9, 12, 13, 14, 15, 16, 18, and 20. The 8 transfer questions were numbers 1, 3, 6, 8, 10, 11, 17, and 19. The 12 sound effects-related questions were numbers 1, 2, 5, 6, 8, 9, 11, 12, 14, 15, and 19. The 8 non-sound effects-related questions were numbers 3, 4, 7, 10, 13, 16, 18, and 20.
**Data Collection**

Seven separate experiment sessions were scheduled in the same computer laboratory classroom to accommodate each participating course section. As participants entered the testing site, they were instructed to sit at the first available computer starting on the left side of the first available row, working to the right. When a row was fully occupied, participants were instructed to begin sitting in the next row in the same manner. By seating participants this way, each version of the narration would be listened to by an equivalent number of participants. When all participants were seated, the Principal Investigator passed out Institutional Review Board-approved consent forms, which the students read and signed. A copy of the consent form is included as Appendix A. The Principal Investigator instructed participants on how to access the narrations by demonstrating from his computer screen, which was projected to the participants. They were then told to listen to the narration only once without pausing, and to take the test as soon as they were finished listening to the story. They were also instructed to circle only one answer per question by using the pen that was provided. The Principal Investigator told each participant that the time limit for the test was 15 min. The pen-and-paper test was turned face down in front of each participant’s computer until the narration was over. The Principal Investigator gave the cue for all participants to click the “Play” button on the media player at the same time and begin listening. The Principal Investigator provided the identical model of consumer-grade headphones and the same model of computer. The audio was also normalized to the same volume level. None of the participants reported having any hearing disabilities, and none reported any problems with the volume of the podcasts. The participants were asked to remain in the room until all participants were finished. They were asked if they had any final questions or concerns. None of them had any such issues, so they were dismissed.
In order to insure that participants stayed on task, the Principal Investigator implemented two methods of observation. First, he monitored the participants’ computer screens by occasionally walking to the back of the room inconspicuously and observing the activity. Second, he made sure to watch when each participant finished listening to the narration. Because all participants started listening at the same time and were instructed not to pause, rewind, or fast forward the narration, all participants would be expected to finish within a few seconds of each other. All of the participants in each session stayed on task and completed the experiment as it was designed.

The test sheets were collected as the participants left the room for each session. The test sheets were then separated into different folders based on the treatment code. The Principal Investigator hand-graded the tests by checking each one against a circled answer key, which was consistent with the participants’ test sheets. The participants completed all of the tests correctly. There were no blank answers, no questions with multiple answers, and all answered were circled clearly. Scores were tabulated based on the raw number of correct answers instead of a percentage. For example, a score of 17 meant the participant answered 17 of 20 questions correctly. The test scores were entered into SPSS 17.0. The results for each group included a mean, standard deviation, $p$ value for significance, and $F$ ratio to measure the differences of means between and within each of the groups. To verify the grading accuracy, a person not connected with the current study graded a total of 16 tests (4 from each group), which accounted for slightly more than 11% of the tests. The 16 test scores were the same as the Principal Investigator’s graded test scores.
**Data Analysis**

The Principal Investigator ran a one-way ANOVA in SPSS Version 17 because the treatment was the only independent variable that will be applied to measure the dependent variable in this study. This determined if at least one mean was significantly different from the others. The high values of $p$ and low values of $F$ pre-empted the need to run post hoc tests such as Tukey’s and Scheffe to determine which treatment group means were significantly different. Post-hoc tests were run to determine if there were gender or age differences within each treatment group.

**Independent variables.** There were four independent variables in the current study. Each of them was directly related to the audio treatment of the narration for each group. The independent variable for the control group was Voice (V). The independent variables for the treatment groups were: Voice and Music (VM), Voice and Sound Effects (VS), and Voice, Music, and Sound Effects (VMS).

**Dependent variable.** There was one dependent variable in the current study, Test Score, which was the raw test score. Each participant’s score was reported as a positive integer from 0-20.
Chapter IV: Findings

The purpose of the current study was to determine if the presence of different combinations of music and sound effects in an audio-only educational narration significantly increased or decreased learning by measuring participants’ retention and transfer test scores. This chapter provides both descriptive statistics results as well as results from an ANOVA analysis for each of the research questions.

The descriptive statistics results for all questions are presented in Table 1. Participants in the “music” treatment group ($M = 18.53$, $SD = 1.38$) scored higher than the other groups. The “voice only” control group ($M = 18.31$, $SD = 1.75$), which was expected to score highest, was the second highest-scoring group, followed by the “music and sound effects” treatment group ($M = 18.28$, $SD = 1.82$). The “sound effects” treatment group ($M = 18.17$, $SD = 2.04$) scored lowest. The results were analyzed by running a one-way ANOVA, which indicated there was no main effect for treatment, $F(3, 138) = 0.271, p = 0.846$. Therefore, the current study failed to reject the null hypothesis.
Table 1

Descriptive Statistics Results for Retention and Transfer Tests in Audio Narration Using Different Treatments (All Questions)

\( V = \text{Voice}; \ VM = \text{Voice, Music}; \ VS = \text{Voice, Sound Effects}; \ VMS = \text{Voice, Music, Sound Effects} \)

<table>
<thead>
<tr>
<th>Measure</th>
<th>V</th>
<th>VM</th>
<th>VS</th>
<th>VMS</th>
</tr>
</thead>
<tbody>
<tr>
<td>M *</td>
<td>18.31</td>
<td>18.53</td>
<td>18.17</td>
<td>18.26</td>
</tr>
<tr>
<td>SD</td>
<td>1.75</td>
<td>1.38</td>
<td>2.04</td>
<td>1.82</td>
</tr>
<tr>
<td>% Correct</td>
<td>91.55</td>
<td>92.65</td>
<td>90.85</td>
<td>91.30</td>
</tr>
</tbody>
</table>

*Note. *\( M \) is expressed as a raw score based on 1 point per correct answer.

The test scores were also analyzed by comparing results separately on retention questions and transfer questions. Twelve retention questions asked the participants to recall facts that were in the content of the story. Results for the retention questions are presented in Table 2. Participants in the “music” treatment group (\( M = 11.33, SD = 0.86 \)) scored higher than the other groups. The “music and sound effects” treatment group (\( M = 11.23, SD = 1.37 \)), scored second highest. The “voice only” control group (\( M = 11.14, SD = 1.15 \)), which was expected to score highest, was tied for third highest, along with the “sound effects” treatment group (\( M = 11.14, SD = 1.10 \)). The results were analyzed by running a one-way ANOVA, which indicated there was no main effect for treatment, \( F(3, 138) = 0.229, p = 0.876 \).
Table 2

Results for Retention and Transfer Tests in Audio Narration Using Different Treatments

(Retention Questions)

\( V = \text{Voice}; \ \text{VM} = \text{Voice, Music}; \ \text{VS} = \text{Voice, Sound Effects}; \ \text{VMS} = \text{Voice, Music, Sound Effects} \)

<table>
<thead>
<tr>
<th>Measure</th>
<th>V</th>
<th>VM</th>
<th>VS</th>
<th>VMS</th>
</tr>
</thead>
<tbody>
<tr>
<td>( M ) *</td>
<td>11.14</td>
<td>11.33</td>
<td>11.14</td>
<td>11.23</td>
</tr>
<tr>
<td>SD</td>
<td>1.15</td>
<td>0.86</td>
<td>1.10</td>
<td>1.37</td>
</tr>
<tr>
<td>% Correct</td>
<td>92.83</td>
<td>94.42</td>
<td>92.83</td>
<td>93.58</td>
</tr>
</tbody>
</table>

*Note. *\( M \) is expressed as a raw score based on 1 point per correct answer.

Eight transfer questions asked the participants to apply critical thinking and problem solving skills to content in the narration. Results for the transfer questions are presented in Table 3. Participants in the “music” treatment group (\( M = 7.19, SD = 0.79 \)) scored higher than the other groups. The “voice only” control group (\( M = 7.17, SD = 0.91 \)), which was expected to score highest, scored second highest. The “music and sound effects” treatment group (\( M = 7.03, SD = 0.75 \)) was tied for the third highest-scoring group, along with the “sound effects” treatment group (\( M = 7.03, SD = 1.18 \)). The results were analyzed by running a one-way ANOVA, which indicated there was no main effect for treatment, \( F(3, 138) = 0.315, p = 0.814 \).
Table 3

Results for Retention and Transfer Tests in Audio Narration Using Different Treatments

(Transfer Questions)

\( V = \text{Voice}; \ VM = \text{Voice, Music}; \ VS = \text{Voice, Sound Effects}; \ VMS = \text{Voice, Music, Sound Effects} \)

<table>
<thead>
<tr>
<th>Measure</th>
<th>V</th>
<th>VM</th>
<th>VS</th>
<th>VMS</th>
</tr>
</thead>
<tbody>
<tr>
<td>( M^* )</td>
<td>7.17</td>
<td>7.19</td>
<td>7.03</td>
<td>7.03</td>
</tr>
<tr>
<td>( SD )</td>
<td>0.91</td>
<td>0.79</td>
<td>1.18</td>
<td>0.75</td>
</tr>
<tr>
<td>% Correct</td>
<td>89.63</td>
<td>89.88</td>
<td>87.88</td>
<td>87.88</td>
</tr>
</tbody>
</table>

Note. \( *M \) is expressed as a raw score based on 1 point per correct answer.

The test scores were also analyzed by comparing results separately on questions that asked about content that was placed immediately after a sound effect was played, and questions about content that was not placed immediately after a sound effect was played. Twelve questions were about content that was placed immediately after a sound effect was played. Results for the sound effects-related questions are presented in Table 4. Participants in the “music” treatment group (\( M = 11.25, SD = 0.91 \)) scored higher than the other groups. The “voice only” control group (\( M = 11.17, SD = 1.13 \)), which was expected to score highest, scored second highest. The “music and sound effects” treatment group (\( M = 11.14, SD = 1.33 \)), scored third highest. The “sound effects” treatment group (\( M = 10.97, SD = 1.38 \)) scored lowest. The results were analyzed by running a one-way ANOVA, which indicated there was no main effect for treatment, \( F(3, 138) = 0.343, p = 0.794 \).
Table 4

Results for Retention and Transfer Tests in Audio Narration Using Different Treatments
(Sound Effects-Related Questions)

\( V = \text{Voice}; \ VM = \text{Voice, Music}; \ VS = \text{Voice, Sound Effects}; \ VMS = \text{Voice, Music, Sound Effects} \)

<table>
<thead>
<tr>
<th>Measure</th>
<th>V</th>
<th>VM</th>
<th>VS</th>
<th>VMS</th>
</tr>
</thead>
<tbody>
<tr>
<td>( M^* )</td>
<td>11.17</td>
<td>11.25</td>
<td>10.97</td>
<td>11.14</td>
</tr>
<tr>
<td>( SD )</td>
<td>1.13</td>
<td>0.94</td>
<td>1.38</td>
<td>1.33</td>
</tr>
<tr>
<td>% Correct</td>
<td>93.08</td>
<td>93.75</td>
<td>91.42</td>
<td>92.83</td>
</tr>
</tbody>
</table>

Note. *\( M \) is expressed as a raw score based on 1 point per correct answer.

Eight questions were about content that was not placed immediately after a sound effect was played. Results for the non-sound effects-related questions are presented in Table 5.

Participants in the “music” treatment group (\( M = 7.28, SD = 0.85 \)) scored higher than the other groups. The “sound effects” treatment group (\( M = 7.19, SD = 1.95 \)) scored second highest. The “voice only” control group (\( M = 7.14, SD = 0.90 \)), which was expected to score highest, scored third highest. The “music and sound effects” treatment group (\( M = 7.11, SD = 0.96 \)) scored lowest. The results were analyzed by running a one-way ANOVA, which indicated there was no main effect for treatment, \( F(3, 138) = 0.125, p = 0.945 \).
Table 5

Results for Retention and Transfer Tests in Audio Narration Using Different Treatments
(Non-Sound Effects-Related Questions)

\(V = \text{Voice}; \ VM = \text{Voice, Music}; \ VS = \text{Voice, Sound Effects}; \ VMS = \text{Voice, Music, Sound Effects}\)

<table>
<thead>
<tr>
<th>Measure</th>
<th>V</th>
<th>VM</th>
<th>VS</th>
<th>VMS</th>
</tr>
</thead>
<tbody>
<tr>
<td>(M^*)</td>
<td>7.14</td>
<td>7.28</td>
<td>7.19</td>
<td>7.11</td>
</tr>
<tr>
<td>(SD)</td>
<td>0.90</td>
<td>0.85</td>
<td>1.95</td>
<td>0.96</td>
</tr>
<tr>
<td>% Correct</td>
<td>89.25</td>
<td>91.00</td>
<td>89.88</td>
<td>88.88</td>
</tr>
</tbody>
</table>

Note. *\(M\) is expressed as a raw score based on 1 point per correct answer.

Although the current study was not intended measure gender or age differences, the results were further analyzed to ascertain whether the findings were consistent among the subgroups of participants. In the control group, there were 20 females (\(M = 18.20, SD = 1.47\)) and 15 males (\(M = 18.73, SD = 1.79\)). The results were analyzed by running a one-way ANOVA, which indicated there was no significant effect for gender, \(F(1, 33) = 0.925, p = 0.343\). In the voice and music group, there were 24 females (\(M = 18.42, SD = 1.32\)) and 12 males (\(M = 18.75, SD = 1.54\)). The results were analyzed by running a one-way ANOVA, which indicated there was no significant effect for gender, \(F(1, 34) = 0.448, p = 0.508\). In the the voice and sound effects group, there were 19 females (\(M = 18.63, SD = 1.38\)) and 17 males (\(M = 17.94, SD = 2.36\)). The results were analyzed by running a one-way ANOVA, which indicated there was no significant effect for gender, \(F(1, 34) = 1.177, p = 0.286\). In the voice, music, and sound effects group, there were 21 females (\(M = 18.29, SD = 1.27\)) and 14 males (\(M = 18.21, SD = 2.49\)). The
results were analyzed by running a one-way ANOVA, which indicated there was no significant effect for gender, $F(1, 33) = 0.016, p = 0.901$.

The test scores for traditional college students, defined as 18-24 years old, and non-traditional college students aged 25-and-over within each group were also compared. In the control group, there were 25 traditional college students ($M = 18.76, SD = 1.51$) and 10 non-traditional college students ($M = 17.60, SD = 1.65$). The results were analyzed by running a one-way ANOVA, which indicated there was no significant effect for age, $F(1, 33) = 4.004, p = 0.054$. In the voice and music group, there were 17 traditional college students ($M = 18.88, SD = 1.32$) and 19 non-traditional college students ($M = 18.21, SD = 1.40$). The results were analyzed by running a one-way ANOVA, which indicated there was no significant effect for age, $F(1, 34) = 2.168, p = 0.150$. In the voice and sound effects group, there were 22 traditional college students ($M = 18.59, SD = 1.97$) and 14 non-traditional college students ($M = 17.86, SD = 1.79$). The results were analyzed by running a one-way ANOVA, which indicated there was no significant effect for age, $F(1, 34) = 1.259, p = 0.270$. In the voice, music, and sound effects group, there were 23 traditional college students ($M = 18.57, SD = 2.04$) and 12 non-traditional college students ($M = 17.67, SD = 1.15$). The results were analyzed by running a one-way ANOVA, which indicated there was no significant effect for age, $F(1, 33) = 1.987, p = 0.168$.

Although no differences emerged in these data, the small sample sizes for each gender and age group were too small to propose any conclusions based on the results. Larger samples of each population would yield more conclusive results about gender and age differences.

During the trials, the Principal Investigator did not observe any noticeable listening fatigue. None of the participants looked to be sleeping, drowsy, disinterested, or distracted. The
Principal Investigator concluded that none of the lower scores were caused by any listening fatigue factors.

**Summary of Results**

The research hypotheses were:

- \( H_1 \): Learning is improved when sound effects are excluded from an audio-only educational narration.
- \( H_2 \): Learning is improved when music is excluded from an audio-only educational narration.
- \( H_3 \): Learning is improved when music and sound effects are excluded from an audio-only educational narration.

The results of this study rejected each research hypothesis.

The null hypothesis for this study was:

- \( H_0 \): Learning is not improved when music and/or sound effects are excluded in an audio-only educational narration.

The results of this study failed to reject the null hypothesis.
Chapter V: Discussion

The purpose of the current study was to determine if the presence of different combinations of music and sound effects in an audio-only educational narration significantly decreased learning. The results of the current study indicated that the use of different sound treatments applied to an audio-only narration did not have a significant effect on retention and transfer scores. In each of the test question analyses, the high values of $p$ and low values of $F$ failed to reject the null hypothesis. The results also showed some consistent trends across the different analyses. In each of the five analyses, the mean was slightly, but not significantly, higher in the “music” treatment group. The “sound effects” treatment group scored lowest or tied for the lowest score in four analyses. The sample means for the control and treatment groups were converted to “percentage correct” scores for the purposes of simulating a realistic grading scale and providing a realistic education context. In each analysis, all of the converted percentage scores of the sample means were in the 87-95% range, and would most likely have been in the same letter grade range of ‘A’ or ‘B’ if the tests were administered as part of a course that employs a standard five-letter grading scale.

The results supported some of the findings of Moreno and Mayer (2000), who found no significant differences among the four treatments in their analysis of scores of matching test questions. The current study was also consistent with the finding of Moreno and Mayer (2000) that there was no significant difference between the control group and the voice plus sound effects treatment group. However, the current study found that, although the difference was not significant, the voice and music group scored highest in all five test question analyses. Mayer and Moreno (2000) found that the voice and music scored significantly lower than the control and voice and sound effects group.
The current study’s results were consistent with the results of the pilot study comparing test scores of participants listening to a voice-only narration vs. a narration enhanced with sound effects (Curran, 2011). The pilot study employed the same five analysis methods as the current study, and there was no significant difference between the means of each group. The five analysis methods were: all questions, retention questions, transfer questions, sound effects related questions, and non-sound effects related questions. As in the current study, the statistical values in the pilot study failed to reject the null hypothesis.

The boundary conditions regarding adults and emotional content could partly explain the difference in the results of the current study when compared with Mayer’s Coherence Principle 2 experiments. Mayer (2009) worked with K-12 student participants who were learning scientific content. Additionally, this study used an audio-only narration instead of an animation. Curran (2011) observed that music might be more of a distraction than sound effects, but the current study refuted that notion. The Principal Investigator did not observe any noticeable listening fatigue among the participants, which is a reasonable expectation for a 6 min narration. The arousal factor inherent to sound effects and music might not have made any difference because the listeners did not need them to remain attentive. In longer narrations, listening fatigue is more likely to occur, so it would be worthwhile to see if sound effects and music decrease learning with longer narrations.

The different nature of the narrated content could be another possible explanation for the dissimilar results of the current study when compared with the Coherence Principle 2 studies. The current study used content related to American Literature, while the Coherence Principle 2 studies contained scientific content. Unfortunately, little literature exists on the use of audio in the English disciplines, so the findings of this study can be used as a basis for more empirical
research in American Literature and other English disciplines. Hew (2009) found that only 10% of English instructors use audio as part of their course learning materials, compared with over 33% of science and information technology instructors. This could be the result of English instructors not being familiar with the associated technologies or not having the necessary production skills. The usage level in the English discipline could be increased if they had access to more research.

**Comparisons with Pilot Study**

Even with the aforementioned research design parameters in place, it is important to recognize and address some external factors that can influence the results. In the pilot study, Curran (2011) suggested that future replications be held at various times of day. The current study was conducted in multiple sessions from early morning until late afternoon. The results suggest that time of day did not affect learning. Other potential deterrents to learning that might have been present in the pilot study did not appear to be present in the current study. This could be explained by the higher test scores in the current study. The post-test in both studies contained 20 questions that were listed in the same order that the material was introduced in the narration. In the pilot study, the means for all questions ranged from 13.17-13.35 (Curran, 2011). In the current study, the means for all questions ranged from 18.17-18.55. The narration in the pilot study, which was centered around Native American lore, was more abstract than the narration in the current study.

The formal lab setting did not appear to intimidate some students or make them uncomfortable. The equipment did not hinder their ability to hear the content. Delineating results by learning style could help predict which types of learners can benefit from audio narrations (Hodges, Stackpole-Hodges, & Cox, 2008). The Principal Investigator did not ask
participants to indicate their learning styles or learning disabilities, but the high test scores indicated that participants with a variety of learning styles were able to successfully process the audio content.

**Limitations**

This discussion would not be complete without acknowledging that studies conducted in a laboratory setting can be rendered irrelevant if they are not adapted to real-life environments. Audio research is a prime candidate for future research in natural settings. Whether they are listening to the radio or to a recorded educational podcast, listeners tend to “multitask”, which means they are attending to other matters (Bolls & Muehling, 2007). This behavior is more apparent with the younger “Millenial Generation” of contemporary college students (Frand, 2000), and multitasking might help foster shorter attention spans (Murray, 2004). Listeners might also “tune out” certain content. For example, drivers listening to the radio have to pay attention to traffic, or they might turn the volume down to talk to a passenger. Students listening to an instructor’s podcast might be answering text messages on their phone or surfing the Web on their laptop computer. These distractions are external competitors for resources, attention and memory, and they can further diminish the effect of audio production elements. Another limitation of the current study was the absence of engagement measures, which would have been another indicator of how well the podcast was produced and how well learning was facilitated. Listener satisfaction and entertainment were not measured, either.

**Implications for Practice**

As previously cited, the dramatic increase in online and hybrid courses has fueled a demand for recorded media content that students can download or stream. This demand has also been boosted because of increased use of media and technology in traditional face-to-face
courses. This demand has forced instructors, some who are inexperienced media producers, to consider producing their own content. The current study helps to answer some general questions about how to construct a media production strategy. The results indicate that instructors can start out simply by recording voice-only narrations with no detrimental learning effects. At the same time, the results indicated that instructors can also incorporate more complex audio elements if they wish, also with no detrimental learning effects. The results suggested that instructors can be flexible and customize their audio production strategies that are commensurate with their skills. For example, an instructor in a non-media discipline who has little or no audio production experience would be advised to producing voice-only narrations, but an audio production instructor can add multiple sound effects and music with numerous edits. As it was discussed in Chapter I, some institutions do not have formal media production training programs for their faculty, and some faculty are not highly-skilled media producers, so the results of this dissertation study might help to answer some questions that instructors and designers have about how to produce effective audio narrations.

The development of Multimedia Learning Theory was a significant event in the instructional design field for a number of reasons. MLT recognized that the use of technology in itself was not an effective way to teach. Mayer (2009) applied research-based theory to multimedia instruction, and he encouraged other researchers to do the same. Thus, MLT research focused on learner-centered principles rather than a teacher-centric approach. MLT research prompted educators to examine specific elements of each multimedia presentation from the learner’s perspective so that instructional would be easier for learners to process. The current study further illustrated how MLT concepts can be adapted and applied to different experimental conditions. It is true that there always be limitations because the number of audio narrations and
the number of variables that can be tested is infinite. However, the current study and MLT can be used as a basis for further research in different disciplines. For example, a researcher can replicate the current study with a different short story or replicate MLT with a different science topic. As more studies are conducted within disciplines, more definite conclusions can be made about the findings, which will help to add more clarity to the body of research, which will give instructors and instructional designers a better path to follow when making decisions about producing audio.

**Implications for Future Research**

Audio research studies that were conducted in a controlled laboratory setting have certainly helped isolate factors that can explain how audio is processed. Paisley (1984) and Lang (1994) believe that a multivariate approach that analyzes audio listening in different environments would be more useful to instructional designers and instructors. Some of the participant-related variables that can be tested include the number of times a participant listened, the playback technology a participant used, the external environment where a participant listened, whether or not a participant was multitasking, and a participant’s level of aptitude in the narration topic. Some narrator-related variables can also be tested, including rate of speech, gender, and voice quality. Additionally, sound effects-related variables can be tested, such as volume, number, and intervals. Future studies on the use of audio and learning materials can continue to yield useful results when effective research methods are replicated in real classroom settings.

The current study’s research design omitted potential confounds in order to limit the scope of the study. None of the participants were familiar with the narrated story, which eliminated any chance of existing knowledge skewing the test scores. Some of the other
conditions set forth in this study purposely did not emulate a natural learning environment. Natural conditions could be gradually implemented in future iterations of this research. The Principal Investigator’s on-site presence insured that the design was carried out as it was intended. By restricting the participants to listening once, the test scores did not need to be analyzed separately by number of listens. Technology-associated confounds were removed by providing uniform computer technology, volume settings, and headphones. In a realistic learning environment, a multitude of confounds will naturally be present, so testing confounds as independent variables in multivariate research trials would be beneficial.

One option for future research is to modify the current study by adding another treatment group that would read a text version of the narrated story to evaluate the effectiveness of audio vs. text. This would replicate the study conducted by Kalyuga, Chandler, and Sweller (1997), but in an educational learning situation instead of an employment training module. Another way to modify the current study is to use a repeated measures design. Each group would listen to four narrations, each of which would be a different story. All participants would listen to the same four stories, but the treatments would be rotated. For example, Group 1 might hear a voice-only version of Story 1, while Group 2 would hear a voice and music version of Story 1, and so forth. This technique would be similar to a radio commercial study (Bolls, Lang, & Potter, 2001). By exposing every participant to each treatment, the study’s validity would be improved and the variability will be lowered.

The Principal Investigator chose the music for the current study without pre-screening it with a focus group of the participants’ peers. In the future, this task could be done by having a focus group listen to a few music samples. The potential benefit would be increased listener interest and engagement with the content. The current study did not measure participants’
enjoyment of the content, music or sound effects, nor did it measure their opinions of the production values, such as volume or realism of the sound effects. By doing this, future research could add more insight to the effective assembly of specific production attributes.

A more thorough analysis of sound effects would be another strategy to continue and strengthen future research. For example, “cueing” sound effects could be added. These sound effects are meant to alert listeners to the introduction of content, but are not directly related to the content. For example, the sound of a bell could be used to cue the participants that important content will follow, but the content is not related to a bell ringing. Researchers could compare the effectiveness of the cueing sound effects vs. the relevant sound effects. More attention should be paid to the intervals of time between the introduction of a sound effect and the presentation of related content to help listeners become oriented to the sound effect.

Specifically, techniques that were used in some of the studies cited in the literature review could be incorporated into future variations of the current study. An analysis of the studies revealed that a variety of dependent variables were used to measure the effectiveness of audio in different settings, and they are not necessarily related. Listener engagement can be enhanced by using music or sound effects (Bolls, 2002; Bolls & Lang, 2003; Bolls, Lang, & Potter, 2001; Bolls & Muehling, 2007; Carnahan, Basham, & Musti-Rao, 2009; Lang, 1994; Potter & Choi, 2006). Although there is some evidence that engagement can improve memory (Potter & Choi, 2006), there is no conclusive evidence that engagement has a positive correlation with learning. Memory and learning are fundamentally different. Memory relates to one’s ability to remember facts, but learning includes transfer skills such as critical thinking, problem solving, and kinetic activities. Audio studies that measure engagement and learning could test the strength of their relationship. The literature review also found that engagement can be measured by
observation (Carnahan, Basham, & Musti-Rao, 2009) or a physiological measurement (Potter & Choi, 2006). The heart, eye, and skin measurements used in the physiological radio commercial studies could be applied to educational audio studies to test possible connections between engagement and learning.

Finally, an audio replication of Mayer’s Coherence Principle 2 experiments would be appropriate for future research. If a researcher can acquire copies of Mayer’s lighting and car brake animations, the audio portion could be extracted from each treatment of the animation. Such a replication would improve the current study’s reliability because it would be using the same content that Mayer used. Then, if the results were similar to the current study’s results, the observation about the visual channel being a hindrance to processing sounds would be stronger.
References


Bolls, P. D. (2002). I can hear you, but can I see you? The use of visual cognition during high-imagery radio advertisements. *Communication Research, 29*, 537-563.


Appendix A

IRB-Approved Recruiting Announcement

“A research study that will examine the effectiveness of different audio presentations is being conducted by a doctoral candidate in the UC School of Education. Participants will listen to a brief audio narration and will answer 20 questions about the content of the narration. The time commitment will be 15-30 minutes, and the research activities will take place in a location on the UC-Clermont College campus. A variety of days and times will be offered to participants to conduct the research activities. An incentive of 5 points extra credit will be given by your instructor if you choose to participate. There is absolutely no obligation to participate, and no penalty if you choose not to participate. Your instructor will offer you an alternative assignment for 5 points extra credit if you choose not to participate. If you would like to volunteer as a participant, or if you need more information, please call the Principal Investigator, Andy Curran at 513-732-8971 or email him: andy.curran@uc.edu”
Appendix B

IRB-Approved Consent Form

Adult Consent Form for Research

University of Cincinnati

Department: Curriculum & Instruction

Principal Investigator: Andy Curran

Faculty Advisor: Maya Israel

Title of Study: The Effect of Adding Relevant Sounds to an Audio-Only Narration: A Three-Treatment Application of Mayer’s Coherence Principle

Introduction:

You are being asked to take part in a research study. Please read this paper carefully and ask questions about anything that you do not understand.

Who is doing this research study?

The person in charge of this research study is Andy Curran of the University of Cincinnati (UC) Division of Teacher Education. My advisor, Dr. Maya Israel is guiding this research.

What is the purpose of this research study?

The purpose of this research is to investigate design variables in producing podcasts.

Who will be in this research study?

Between 120-160 people will take part in this study. You may be in this study if you are an adult student (18+ years of age) who is enrolled at UC Clermont College during Winter or Spring Quarter 2012. Due to the nature of the research, individuals with sensory disabilities relative to hearing that are not aided with mechanical or digital devices will not be eligible.

What if you are an employee where the research study is done?

Taking part in this research study is not part of your job. Refusing to be in the study will not affect your job. You will not be offered any special work-related benefits if you take part in this study.
What will you be asked to do in this research study, and how long will it take?

You will be asked to listen to a short (six (6)-minute) audio podcast and answer 20 questions about the content. It will take roughly 20 minutes to listen and take the test. All research activities in this study are experimental. The research will take place in a University of Cincinnati classroom within Clermont College.

Are there any risks to being in this research study?

Possible risks from being in this research study include the following:

- A minimal risk of test anxiety
If you indicate discomfort during the research, you will be excused.

Are there any benefits from being in this research study?

You likely not get any benefit because of being in this study. But, being in this study will help instructors and instructional designers understand how students process and react to different elements of audio.

Will you have to pay anything to be in this research study?

You will not have to pay anything to take part in this study.

What will you get because of being in this research study?

If agreed upon by your instructor, you could receive extra class credit by the instructor of the course from which you were recruited. The amount of extra credit will be determined by the course instructor and will be disclosed during the recruiting process.

Do you have choices about taking part in this research study?

If you do not want to take part in this research study you may quit at any time without penalty. You will not be treated any differently for not participating.

How will your research information be kept confidential?

Your information (written and digital) will be kept in a locked cabinet in the PI’s campus office for three (3) years after the study closes. Signed consent documents will be kept in a separate locked cabinet in the PI’s campus office for four (4) years after the study closes. All information and signed consent forms will be destroyed by electric shredder. Identifiers such as name, birthdate, and UC M-number ID will be deleted before the documents are initially stored. Computerized files related to this study will be deleted three (3) years after the study closes. The data from this research study may be published; but you will not be identified by name.

Agents of the University of Cincinnati may inspect study records for audit or quality assurance purposes.
What are your legal rights in this research study?

Nothing in this consent form waives any legal rights you may have. This consent form also does not release the investigator, the institution, or its agents from liability for negligence.

What if you have questions about this research study?

If you have any questions or concerns about this research study, you should contact Andy Curran: andy.curran@uc.edu or 513-732-8971.

The UC Institutional Review Board (IRB) reviews all research projects that involve human participants to be sure the rights and welfare of participants are protected.

If you have questions about your rights as a participant or complaints about the study, you may contact the Chairperson of the UC IRB at (513) 558-5259. Or, you may call the UC Research Compliance Hotline at (800) 889-1547, or write to the IRB, 300 University Hall, ML 0567, 51 Goodman Drive, Cincinnati, OH 45221-0567, or email the IRB office at irb@ucmail.uc.edu.

Do you HAVE to take part in this research study?

No one has to be in this research study. Refusing to take part will NOT cause any penalty or loss of benefits that you would otherwise have.

You may start and then change your mind and stop at any time. To stop being in the study, you should tell Andy Curran at andy.curran@uc.edu or 513-732-8971.

Agreement:

I have read this information and have received answers to any questions I asked. I give my consent to participate in this research study. I will receive a copy of this signed and dated consent form to keep.

Participant Name (please print) ____________________________________________

Participant Signature ____________________________________________ Date ________
APPENDIX C

Story Narration Script

Title: City Bus

Adapted from a short story by Anonymous

(MUSIC UP AND UNDER)

The road is almost at a stand still. It’s peak time for commuters going home after the work day. A double-decker bus inches along the road. It is full of travellers.

(SFX: BUS SOUND EFFECTS)

Sophie sits at the front of the top deck. Her feet rest by the window, her legs bent. She has earphones in, listening to a jazz song on her iPod.

(SFX: SONG PLAYING THROUGH SPEAKER)

Behind her all the seats are taken. A dour-looking man sporting a heavy beard and wearing a white turban on his head sits immediately behind her. His bag rests on his knees. When Sophie saw the man walk back to his seat, she felt a little uncomfortable.

Sophie looks down at smartphone in her hand. She scrolls through the pictures. They show her and Joe in various poses. Some romantic, some funny…but they are clearly boyfriend & girlfriend. Joe was two days into a week-long business trip, and they kept in touch often with phone calls and text messages.

She looks out of the window at the busy shops. Her reflection shows a frown on her face. She looks back to her smartphone and scrolls through the options...

(SFX: PHONE BEEPS-BOOPS)

She gets to the NEW MESSAGE option and keys in: JOE I STILL LOVE U. I MISS U. Sophie

(SFX: KEY PAD NOISE)

Sophie looks out the window again, deep in thought about Joe. The man behind her looks nervous. He sweats and fumbles around in his bag, trying to find something.

The man’s search was so frantic that Sophie could hear him even through the sound of music in her earbud headsets.
Sophie turned around to look at the man, who glared back at her. She wondered if she should alert the bus driver to this suspicious behavior, but hesitated. She turned to look ahead of her, biting her bottom lip, nervous. In an attempt to block the noise the man was making, she pulled her I-POD out of her pocket and turns the sound up.

She put it back in her pocket and rested her head back, closing her eyes, and drifting off to sleep for a short cat nap.

Fifteen minutes later, her eyes opened. She blinked, yawned and looked around, quickly scanning the other bus riders. When she turned around, she noticed that the seat behind her was empty, but the man’s bag was still there, under the seat. She looked around the bus to see if he changed seats, but she didn’t see him. She figured that he got off the bus when it stopped during her nap.

She wondered if the man forgot his bag or left it on purpose. Based on his dress, demeanor, and odd behavior, she concluded that he was up to no good and left a bomb or some other explosive device in the bag. Panicking, she pulled the overhead cable to signal the bus driver that she wanted to get off at the next stop, which was two blocks ahead.

As the bus slowed to a stop, Sophie made her way to the front exit door with a few other passengers and began to disembark when the big door opened.

Not wanting to induce panic, Sophie told the driver in a low voice that he should empty the bus and call the police to investigate the suspicious package. He nodded his head and she got off the bus.
(SFX: TRAFFIC NOISE)

She was in a neighborhood that was a 20-minute walk from her home. She took a minute to orient herself so she could figure out which way to walk. Turning north, she glanced back at the corner and noticed the rest of the passengers getting off the bus. She felt relieved that the bus driver took her advice.

Eventually, she got to her apartment and went into the kitchen to fix dinner. Since she was eating alone, she kept it simple: canned vegetable soup and microwave chicken alfredo.

(SFX: MICROWAVE BELL)

She complemented her dinner with a glass of water and ate slowly, telling herself that she’ll cook a real dinner when Joe got back.

(SFX: EATING AND DRINKING SOUNDS)

After Sophie finished, she cleaned up the kitchen.

(SFX: WATER RUNNING)

The dishes were especially messy. She hadn’t washed them in a few days. She was happy for modern conveniences and not having to wash them by hand anymore.

(SFX: DISHWASHER RUNNING)

Sophie had a busy day at work and a semi-traumatic experience on the bus, so felt entitled to relax for the rest of the night. She noticed that it was 5:59, and suddenly realized that the 6:00 news would be starting soon. She wondered if anything came about from the bus incident.

(SFX: TV IN BACKGROUND)

She tuned to Channel 8, which always seemed to scoop the other stations on the breaking news stories.

Sure enough, the lead story had their top reporter, Chris Perdan, on location in front of the bus Sophie rode on that day. Chris recapped the incident in a serious tone, then showed video of a police department bomb squad officer opening a bag that contained…..a crumpled sandwich wrapper and a half-empty bottle of orange juice.

Sophie didn’t know whether to laugh or be embarrassed. She wondered why she was so paranoid about the guy behind her. She questioned her own judgment on reporting the incident to the bus driver. The report concluded with the police chief saying whoever reported this did the right thing, and that it’s better than not reporting something that turned out to be deadly.
The chief’s response gave her some measure of comfort, but she wondered if others had the same perception about the man when they saw him. She also wondered about what happened if the bag was explosive and she kept her mouth shut.

She closed her eyes and imagined how loud the sound would have been, and how many lives would have been ruined. She shuddered and sat on the couch in silence until she fell asleep an hour later.
Appendix D

RETENTION AND TRANSFER TEST

TEST CODE: _____

TEST ID#: _____

Participant information:

Gender: ________ Age: __________ School Class (circle one): FR SO JR SR GR

Program Major: __________________________________

Please circle your answers.

1. What time of day did the story take place?
   a. Morning
   b. Lunchtime
   c. Late Afternoon
   d. Overnight hours

2. Which type of music was Sophie listening to on the bus?
   a. Jazz
   b. Rock
   c. Country
   d. Rap

3. What would be the most logical reason for Sophie to be nervous about the man behind her?
   a. He was a mean man she knew from her old neighborhood.
   b. His clothing seemed to fit the profile of a terrorist.
   c. She was afraid of all men.
   d. Someone sitting next to her warned her about him as he boarded the bus.

4. What was Sophie’s boyfriend’s name?
   a. Mike
   b. Andy
   c. Sal
   d. Joe
5. What was the man’s behavior when he was searching through his bag?
   a. Frantic
   b. Calm
   c. Humorous
   d. Methodical

6. How did Sophie first react to the man searching through his bag?
   a. She immediately told the bus driver
   b. She never even noticed him searching the bag.
   c. She wasn’t sure what to do.
   d. She started yelling out loud about a suspicious package.

7. What did Sophie do when she woke up from her nap on the bus?
   a. She yawned a few times and looked out the window.
   b. She yawned a few times and looked around the bus.
   c. She yawned a few times and put her sweater on.
   d. She yawned a few times and got off the bus.

8. How did Sophie interact with the bus driver when she exited the bus?
   a. She ignored him.
   b. She excitedly told him about the suspicious bag.
   c. She shook him by the shoulders and pleaded with him to empty the bus.
   d. She calmly told him about the suspicious bag.

9. How long would it typically take Sophie to walk home from where she exited the bus?
   a. 10 minutes
   b. 12 minutes
   c. 15 minutes
   d. 20 minutes

10. After Sophie got off the bus, why did she feel relieved?
    a. The bus kept going, so she knew a bomb wouldn’t explode near her.
    b. The bus driver listened to her advice.
    c. She would be home soon.
    d. Her boyfriend would be home in a few days.
11. What is the most accurate way to describe Sophie’s meal?
   a. Gourmet
   b. Take-out
   c. Heat-and-eat
   d. Home cooking

12. What did Sophie drink with her dinner?
   a. Water
   b. Coke
   c. Beer
   d. Wine

13. After cleaning the kitchen, what did Sophie want to do?
   a. Relax
   b. Clean the living room
   c. Call her mom
   d. Go to the movies with her friends

14. What did Sophie watch on TV?
   a. Comedy shows
   b. The news
   c. A game show
   d. A hockey game

15. Which TV channel was she watching?
   a. 2
   b. 4
   c. 6
   d. 8

16. What was in the suspicious bag?
   a. A bomb
   b. A suicide note
   c. A wrapper and a juice bottle
   d. A six-pack of beer
17. What was Sophie’s reaction after the police found the contents of the suspicious bag?
   a. She screamed in horror
   b. She wasn’t sure how to react
   c. She laughed
   d. She was embarrassed

18. What did the police chief say about the person who reported the incident?
   a. The person should never have done it.
   b. The person should have checked the bag before making the report.
   c. The person did the right thing.
   d. The person should have called the police instead of telling the bus driver.

19. What was Sophie’s self-reflection after viewing the chief’s response?
   a. She was angry.
   b. She was scared
   c. She was worried.
   d. She felt comforted.

20. What was the last thing Sophie did in the story?
   a. She fell asleep.
   b. She walked outside.
   c. She called Joe.
   d. She made a cup of tea.