Design Effects of Screen-Captured Tutorials on Student Achievement

JOE WINSLOW, PH.D., JEREMY DICKERSON, ED.D. & CHENG-YUAN LEE, PH.D.
Spadoni College of Education
Coastal Carolina University

ABSTRACT
This paper reports findings from an experimental study of the effects of using differentiated screen-captured multimedia tutorials to teach technology integration skills to pre-service educators. A homogeneous set of participants (N=96) were randomly assigned to one of four treatment groups. Participants in each group completed a web-based instructional module using screen captured tutorials developed as Graphic plus Text, Video, Video plus Narration, and Video plus Text. A post-test was administered following the module completion to assess both conceptual recall and task performance. Participants receiving video plus narration tutorials outperformed those in other groups, particularly those receiving only video. These findings are supported in the literature through the cognitive load principles of modality and redundancy. These results add to the growing body of knowledge about the design and development of multimedia instructional materials and invite further inquiry into related covariates that potentially influence the design of screen captured tutorials.
INTRODUCTION

The integration of multimedia technology continues to reshape the field of teacher education. In the age of mobile devices and web ubiquity, and given the widespread acceptance of distance learning, most universities now offer programs having some combination of face-to-face, blended or fully online courses in order to meet the needs of pre-service and in-service educators. In particular, the movement toward online delivery has been heavily influenced by national legislation for student achievement and teacher accountability, with increased emphasis on cyclical accreditation by national agencies, such as NCATE (National Council for the Accreditation of Teacher Education). As a result, teacher education institutions face growing expectations to offer innovative programs that leverage the enhanced processing power and bandwidth of modern systems to deliver cost-effective, efficient, and high-quality instruction. Among other topics, teacher education involves the study of technology as both subject matter and instructional strategy. Students learn about technology, with technology, as well as the integration of technology for the purposes of developing, designing, and delivering instruction. Such courses are usually referred to as technology integration or instructional technology courses, and they prepare students to learn computer operations and software skills, as well as knowledge about how these skills can be used in the classroom to support student learning, facilitate instructional delivery, and promote citizenship in a technological society.

Online teaching and learning is a process inundated with tools and techniques. Of particular interest in this study was the design and development of screen-captured tutorials (SCTs) as instructional resources for online courses. SCTs are multimedia files captured as animated videos of desktop screen operations, supplemented with text callouts, graphic cues, static pictures, and/or audio narrations to present concepts and techniques relevant to technology instruction, such as software utilization (Drumheller & Lawler, 2011). SCTs may be used independently as standalone learning objects, or embedded into other technologies such as PowerPoint® presentations, PDFs, web sites, social media, mobile apps, and similar. Given the continued increase in online course delivery in teacher education, coupled with advances in the usability of multimedia authoring tools, the effective design and development of screen captured tutorials will continue to be relevant for faculty and student audiences training to teach in face-to-face, blended, or virtual school environments that integrate on-demand multimedia resources. This study presents empirical data and analysis about the differential achievement effects of various screen captured tutorial designs within the framework of multimedia learning theories and cognitive load research.

PREVIOUS INQUIRY

A vast array of technologies can be used to develop learning environments with electronic information in different representational formats, from basic PowerPoint® presentations to advanced interactive virtual realities such as Second Life® (Schnitz & Kürschner, 2007). Multimedia tools have the potential to promote meaningful learning by varying both the type of representations provided to students and the degree of student interactivity (Moreno & Valdez, 2005). The design process is not to be overlooked; research suggests that well-designed multimedia materials encourage the learning process, particularly when there is purposeful effort to not overload working memory and not redirect learner attention from relevant processes (Kester, Kirschner, & van Merriënboer, 2005; Mayer & Johnson, 2008). The learner's ability to generate mental representations of external information is enhanced when incoming information is presented through multiple sensory
channels (Zheng, McAlack, Wilmes, Kohler-Evans, & Williamson, 2009). Multimedia tools have the capacity to integrate different learning modalities, but choosing what to use (and why to use them) can be challenging.

Researchers have developed several theoretical models explaining multimedia learning, including Paivio's dual coding (1986), Gagne and Driscoll's information processing (1988), and Mayer's cognitive load (2003). Cognitive load research is critical in the area of pre-service teacher education because of the demands of classroom teaching and the need for a high level of teacher performance and accountability (Feldon, 2007). Cognitive load refers to the mental load or stress placed on working memory based on intrinsic and extrinsic factors that need to be managed during the learning process (Kenny, 2009). Cognitive load theory (CLT) collectively has roots and connections to teaching and learning theories, biology, neuron research, cognitive brain science, and philosophy (Ayres, & Paas, 2007). Cognitive load is a central consideration in the use of technology tools for learning and the design of multimedia instruction (Mayer, & Moreno, 2003). Intrinsic and extraneous cognitive load in visual displays and information can be manipulated to moderate design effectiveness (Hyunjeong, Plass, & Homer, 2006). According to Seufert & Brünken (2006), the learning of complex content is typically more effective when extraneous cognitive load is reduced to free up resources for primary processes and learning. Research by van Merriënboer, & Ayres, (2005) suggests that CLT is becoming increasingly useful for the design of interactive online courses.

Many major studies have identified mixed or contradicting results investigating multimedia design effectiveness (Rieber, 1990; Park and Hopkins, 1993; Anglin, Vaez, and Cunningham, 2003). According to Zhu and Grabowski (2006), the vast majority of studies comparing the use of static and animated modalities have focused primarily on conceptual representation or explanation. Veronikas and Maushak (2005) found no significant achievement effects among groups receiving software instruction via screen captured images with text, audio, or both. Sometimes image and text animation can hinder rather than improve learning (Hasler, Kersten, & Sweller, 2007). Compared to static pictures, animated pictures provide additional information, which has been noted in the literature to have different effects on learning (Schnotz, & Rasch, 2005). Based on a cognitive processing-based design framework, Nelson, & Erlandson (2008) suggest removing as much extraneous graphical and textual information as possible from the learning environment. Further, Moreno's (2007) research encourages segmenting instructional videos and animations into small chunks to help novice students learn from complex dynamic visualizations. Morrison, & Anglin (2005) discuss the potential for presenting two integrated non-redundant external representations (verbal and visual) in contrast to one (verbal or visual) and suggest that this model typically results in higher student performance levels and requires less mental effort by learners.

Screen captured tutorial design, an emerging category of multimedia learning, has received limited attention in the literature to date. Of the few studies reported, researchers have investigated SCT design primarily for content instruction (Lin & Atkinson, 2011; Walker et al., 2011; Aldalalah, 2010), or pedagogy instruction (Curcher, 2011; Parette et al., 2011). Only two studies were located that examined SCT design within the context of technology instruction, and both of those were focused on online library search systems (Laurie, 2009; Wales & Robertson, 2008). No studies were located investigating the use of SCTs for teaching software applications to educators, which is the focus of this investigation. Thus, a significant rationale for the study reported in this paper is its novelty, especially when interpreted within the context of previous inquiry on general multimedia design for online education.
PROBLEM

This study was conducted within an instructional technology course in a teacher preparation program at a public, mid-sized university in the United States. Course content emphasized basic technology operations, concepts, planning, integration, assessment, and citizenship, aligned to various state and national standards. To facilitate consistency and quality control of the course across multiple sections and faculty each semester, the researcher designed a comprehensive sequence of web-based instructional modules containing text, images, and links to multimedia files and homework documents accessible via the Blackboard course management system.

During early pilot testing, students reported that screen-captured tutorials (SCTs) were the most useful learning objects in modules that focused on software skills development. These tutorials were short videos that demonstrated correct mouse interactions in target software applications with an accompanying voice-over narration. Although the duration of each video rarely exceeded more than a few minutes, the development time necessary to create them for every module in the course was daunting—audio scripts had to be written, voiceovers had to be rehearsed, and captured clips had to be carefully edited in post-production before a final product could be exported. Thus, before committing to this production overhead for the entire course module sequence, the researcher decided that an empirical investigation was necessary to determine if the massive amount of development effort was justified.

Two research questions of this study were:

1. Would students receiving software instruction via screen-captured video tutorials demonstrate different achievement than students receiving instruction via tutorials designed with static screen shots and supporting text?
2. Will adding text cues or voiceover narrations to the screen-recorded video tutorials result in different achievement levels?

PARTICIPANTS

The sample for this study was drawn from the entire population (N=121) of undergraduate teacher preparation students enrolled in six sections of a required instructional technology course. The majority of these students were female (78%) aged 18-24, with the proposed majors of Early Childhood, Elementary, Middle Grades or Physical Education. Participants were tested and surveyed extensively at the beginning of the course to determine prior knowledge and pre-existing technology skills. Data from the instruments indicated strongly that the majority of participants were basic users of the software application targeted in the study (Microsoft PowerPoint©). Although most participants were capable of generating a simple slideshow, only three students were competent in the advanced skills that were the focus of the instructional treatments designed for this study. Of these three, only one was present during the actual experiment, and her data were subsequently purged from the analysis to preserve the homogeneity of the sample. This reduction, combined with absenteeism on the day of the experiment, yielded a final sample size of N=96.
TREATMENTS

The instructional content targeted in this study focused on teaching participants two advanced PowerPoint® skill sets: 1) inserting styled hyperlinks to create non-linear slideshow navigation, and 2) applying custom animation triggers to control progressive content disclosure. The application context for the project document in each treatment was an interactive Jeopardy®-style game about middle school science concepts. Four instructional tutorials were developed to teach identical software task sequences using identical content in different representational formats – Static Graphics with Text (ST), Video (V), Video with Narration (VN), and Video with Text (VT). The tutorials were produced using TechSmith Camtasia® and saved as Flash files (.swf) for playback through a web browser.

Participants were randomly assigned to one of the four treatment groups and briefed with a scripted statement describing the protocol for the study. At the conclusion of the briefing, participants entered a specific URL unique to their treatment group. Each treatment consisted of eight sequential screens that displayed identical summary text plus a graphic frame for the ST, V, VN or VT content. Cueing text in the VT group was identical to the narration track in the VN group. Time and instructional venue were constant for all participants. Two assessments were conducted at the conclusion of the treatments—a posttest for conceptual recall, and a performance analysis of the PowerPoint® file.

FINDINGS

One-way ANOVA was used to analyze the data from both the recall and performance assessments. Results from the recall posttest (Table 1) indicated no significant difference between treatment groups (p<.29). Analysis of the performance assessment data (Table 2) yielded a significant difference between the Video only (V) group and the Video plus Narration (VN) group at p<.0001. The differences between the Static Graphic with Text group (ST) and the Video plus Narration group (VN) were approaching significance at p<.09, as were VN and VT, also at p<.09.

<table>
<thead>
<tr>
<th>TABLE 1. RECALL ASSESSMENT</th>
</tr>
</thead>
<tbody>
<tr>
<td>Groups</td>
</tr>
<tr>
<td>ST - Static Graphic with Text</td>
</tr>
<tr>
<td>V - Video Only</td>
</tr>
<tr>
<td>VN - Video with Narration</td>
</tr>
<tr>
<td>VT - Video with Text</td>
</tr>
</tbody>
</table>

ANOVA

<table>
<thead>
<tr>
<th>Source of Variation</th>
<th>SS</th>
<th>df</th>
<th>MS</th>
<th>F</th>
<th>P-value</th>
<th>F crit</th>
</tr>
</thead>
<tbody>
<tr>
<td>Between Groups</td>
<td>1287.5</td>
<td>3</td>
<td>429.16</td>
<td>1.28</td>
<td>0.29</td>
<td>2.70</td>
</tr>
<tr>
<td>Within Groups</td>
<td>30953.2</td>
<td>92</td>
<td>336.45</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>32240.6</td>
<td>95</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
TABLE 2. PERFORMANCE ASSESSMENT

<table>
<thead>
<tr>
<th>Groups</th>
<th>Count</th>
<th>Sum</th>
<th>MEAN</th>
<th>SD</th>
</tr>
</thead>
<tbody>
<tr>
<td>ST – Static Graphic with Text</td>
<td>25</td>
<td>1726</td>
<td>69.04</td>
<td>32.42</td>
</tr>
<tr>
<td>V – Video Only</td>
<td>26</td>
<td>1364</td>
<td>52.46</td>
<td>35.21</td>
</tr>
<tr>
<td>VN – Video with Narration</td>
<td>21</td>
<td>1881</td>
<td>89.57</td>
<td>16.63</td>
</tr>
<tr>
<td>VT – Video with Text</td>
<td>24</td>
<td>1657</td>
<td>69.04</td>
<td>28.15</td>
</tr>
</tbody>
</table>

ANOVA

<table>
<thead>
<tr>
<th>Source of Variation</th>
<th>SS</th>
<th>df</th>
<th>MS</th>
<th>F</th>
<th>P-value</th>
<th>F crit</th>
</tr>
</thead>
<tbody>
<tr>
<td>Between Groups</td>
<td>15998.3</td>
<td>3</td>
<td>5332.77</td>
<td>6.13</td>
<td>0.001</td>
<td>2.70</td>
</tr>
<tr>
<td>Within Groups</td>
<td>79995.5</td>
<td>92</td>
<td>869.52</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>95993.8</td>
<td>95</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

DISCUSSION

The data from this study seem to support the modality principle, as explained by Mayer (2003). Participants who watched only a screen captured video of correct mouse operations performed significantly worse than those who watched a video with an accompanying narration audio track. The modality principle suggests a complementary relationship can coexist between simultaneous narration and visual media, as opposed to a competitive relationship between narration and text media—sometimes called a split-attention effect (Ayres & Sweller, 2005).

It is not surprising that participants in the Video with Narration (VN) group outperformed those in the Video with Text (VT) group. Cognitive load theory explains this design effect with its supposition that combining media with both high intrinsic and high extrinsic loads (such as reading text and listening to narration) results in exceeding the capacity of working memory, whereas a combination having high intrinsic and low extrinsic load (such as watching a video and listening to the narration) stays within the capacity of working memory (Ayres, 2005).

Given that the text callouts were identical to the narrations and occasionally lengthy, Mayer’s redundancy principle (2003) also suggests that the learners likely spent extra mental effort comparing the printed and spoken text, thereby creating extraneous processing and diminishing student focus. The text callouts also competed visually for learner attention, and may have distracted eye movements away from the software operations being demonstrated onscreen.

CONCLUSIONS AND FUTURE RESEARCH

Given that the amount of effort required to script and record narration adds significantly to the tutorial production workflow, the findings of this study provide some evidence that the extra effort is worthwhile; learner performance improves when narration is added to screen captured video. The findings also suggest to a less significant degree that text callouts (either on static screen shot graphics or embedded screen captured videos) are less effective than the combination of screen captured video and accompanying narration. While this finding supports Mayer’s original redundancy principle (2003), he revised the
principle recently, noting that “short redundant phrases (can) guide the learner’s attention without priming extraneous processing” (2008). Consequently, the authors suggest that future research in screen captured tutorials examine the optimal length of text callout verbiage relative to the narration script.

Mayer’s multimedia theory also argues that learners perform better when their attention is appropriately focused on key elements in a lesson; his signaling principle suggests that embedding verbal cues, such as vocal emphasis, reduces extraneous processing (2009). Mayer suggests that future inquiry is needed to determine the effectiveness of visual cueing, “such as highlighting portions of a graphic that correspond to ongoing narration” (2009). Given that screen captured tutorials rely heavily on a learner’s ability to track mouse movements and interactions within a target software program, the researchers agree that the use of visual signals, such as flash identifiers or colored symbols, should be investigated for any benefit to guiding learner attention.

Previous inquiry has demonstrated that adding “social presence” in a computer-based learning environment can have a positive effect on learner motivation and active cognitive processing (Lepper, Woolverton, Mumme, & Gurtner, 1993; Mayer et al., 2004, Mayer, Sobko, & Mauton, 2003). One such design technique that is consistent with Mayer’s personalization principle (2009) involves the recording of narration in conversational style rather than in third-person, formal style. The researchers suggest that this principle should be investigated in future studies of screen captured tutorial design because software utilization is ultimately a personal process where the learner must physically interact with a machine, and efforts to humanize the experience may have the potential to reduce extraneous processing caused by computer anxiety (Chien, 2008).

Further research is also needed to investigate the optimal granularity of screen captured tutorials. If an average adolescent or adult learner can normally process a task in working memory intently for only a few minutes before experiencing mental fatigue or boredom (Sousa, 2009), then how should screen captured tutorials be segmented to teach long software processes? How many mouse clicks are necessary to record what Mayer calls the correct “bite size” in his segmenting principle (2009)? And to what degree are these effects mediated by learner control? Do learners perform equally well when the screen captured tutorial is a single continuous (non-segmented) file that can be paused and played back at any point? Or should designers segment long captures as a sequence of shorter individual clips?

Research on the design of multimedia learning is still in its early stages. Although the foundation of empirical and theoretical support continues to build, the rapid pace of technological change requires constant and diligent review of emerging designs. As cloud and mobile technologies extend their reach into academic environments, students and teachers will continue to be expected to learn to use new software tools. Screen captured tutorials therefore deserve continued attention in the research community to identify design attributes that improve performance.

Direct Reprint Requests to:
Email: jwinslow@coastal.edu
REFERENCES


