JRTE, *41*(2), 203–221

Effects of Multimedia Software on Achievement of Middle School Students in an American History Class

Karla V. Kingsley
University of New Mexico

Randall Boone
University of Nevada Las Vegas

Abstract

This study investigated social studies achievement as a result of utilizing a multimedia-based American history software program (Ignite Early American History, 2003) to augment textbook and lecture materials for seventh-grade middle school history students in an ethnically and linguistically diverse urban school district. The instructional software used was an interactive multimedia program designed to teach middle school students through video, song, animation, text, and other media to develop critical thinking skills while acquiring knowledge of required content strands (Ignite Learning, 2003). Teacher and student activities, pretest and posttest scores, and instructional methods for experimental and control conditions were documented in order to provide a comprehensive understanding of the results. (Keywords: multimedia instruction, instructional technology, educational software, social studies achievement.)

INTRODUCTION AND PURPOSE OF THE STUDY

While there is a significant body of literature that discusses technology integration in schools and classrooms, there remains a dearth of data-based research specifically addressing the issue of the effectiveness of different types of educational software in relation to student achievement outcomes (Crosier, Cobb, & Wilson, 2002; Mills, 2001; Williams, Boone, & Kingsley, 2004). The National Research Council (2002) and others (Campbell, 1969; Cook, 2001) found repeatedly that although most educational software is commercially produced, "Those with commercial interests are not expected by educators, policy makers or the public to use research to support what they sell" (National Research Council, 2002, p. 96). Consequently, the National Research Council (NRC) explained, "Educators are unlikely to draw on scientific knowledge to improve their practices in any meaningful way" (p. 96). The current culture of high-stakes testing in the United States, with its focus on high expectations and accountability, has sparked renewed interest in identifying effective educational interventions, including computer assisted instruction, that increase student achievement. This investigation adds to the small body of studies utilizing a rigorous, scientifically based research methodology, as defined by the Institute of Education Sciences (IES), to examine student outcomes as a result of a technology intervention within a school setting (Poggi, 2003).

Interactive Media Focus

The term *multimedia* describes any system that combines two or more media into a single product or presentation, such as a software program or a Web page. Although interactive multimedia capabilities are constantly evolving and have become very popular among educators in recent years, the body of research on interactive multimedia as an instructional approach is not yet extensive (Alessi & Trollip, 2001; Lockard & Abrams, 2004). According to Mayer (2003), a multimedia instructional message is "a presentation consisting of words and pictures that is designed to foster meaningful learning. Thus, there are two parts to the definition: (a) the presentation contains words and pictures, and (b) the presentation is designed to foster meaningful learning" (p. 128).

Mayer (2003) and others (Brouwer, Muller, & Rietdijk, 2007; Thompson, 2007) have emphasized the unique contributions multimedia brings to the learning experience. There are data to support the assertion that multimedia capabilities are unique because both sensory stimulation and user navigation in interactive multimedia (IMM) parallel students' natural ways of learning (Bagui, 1998; Gibbs, Graves, & Bernas, 2001). Roblyer (1999) asserted that the multiple channels through which multimedia communicates to the learner seem to be the source of its benefits. The sound, images, animation, and interactivity in electronic books have also been shown to increase motivation and comprehension scores as compared to students' reading of printed texts (Greenlee-Moore & Smith, 1996; Labbo, 2002). According to some researchers (Becker, 2000; Mayer, 2003; Moreno, 2006), interactive multimedia is one of the best technologies to help students learn. Although claims such as this one elicit varying responses among scholars and educators, some research appears to indicate that IMM can indeed provide learning benefits (Hancock, Knezek, & Christensen, 2007).

Media-focused research. Much multimedia research has focused on the specific media employed for instruction. Clark and Mayer (2003) provided a list of several research-based principles for instructional multimedia that focused on media type. For example, their *multimedia* principle discussed the use of text with accompanying graphics as opposed to text alone. Similarly, their *contiguity* principle looked at the proximity and placement of a graphic to its corresponding text. And while much of this research is based on design principles that focus on cognitive processing, especially cognitive load theory (Brünken, Plass, & Leutner, 2003; Mayer, 2005; Moreno & Valdez, 2005; van Merriënboer & Sweller, 2005), priority remains centered on the media elements.

Supportive Resource Focus

While the instructional modalities of the software used in this study included the expected multimedia components, text remained the predominant vehicle for instruction. Text was incorporated into alternate contexts such as maps, matching problems, text-documented illustrations, document facsimiles, timelines, concurrent text with spoken audio, and karaoke-style song lyrics.

With electronic *text transformations* (e.g., timelines, document facsimiles) comprising the bulk of the content of the software used in this study, the Clark and Mayer (2003) media-focused theoretical construct prevalent in much previous research did not provide a satisfactory framework for consideration of this software as it was implemented. Rather, a typology of supportive resources developed by Anderson-Inman and Horney (1998, 2007) that describe how the process of reading can be made easier or more educational (Horney & Anderson-Inman, 1999) provided the framework for a functional matrix of the educational resources in the software.

Several of these text transformations closely resembled well-documented instructional strategies (e.g., advance organizers, graphic organizers, visual displays such as timelines, and mnemonic devices) from research into adapting challenging textbooks for students experiencing difficulty in reading (Boone & Higgins, 2005; Higgins & Boone, 2001; Higgins, Boone, & Lovitt, 2002). The overwhelming reliance on text for content delivery in this software suggested an instructional design relying more on a concept of supported electronic text (Anderson-Inman & Horney, 2007) than the traditional media interactions often associated with multimedia.

Typology of resources. A recently revised description of the Anderson-Inman and Horney (2007) typology of resources for supported electronic text included 11 resource types: presentational, navigational, translational, explanatory, illustrative, summarizing, enrichment, instructional, notational, collaborative, and evaluational. Anderson-Inman and Horney stated that the advantage of a typology that does not "focus on what media is being used to modify or enhance the electronic text, but rather what function the supportive resources play in the reading process" (p. 153) is its usefulness for "teachers, students, and parents…to think critically about the modifications, enhancements, and additions they encounter when selecting or reading electronic versions of assigned texts" (p. 154).

Supported electronic text. Overlaying this framework on the supportive resources from the American history software used in this research, a clearer design for instruction emerged. Resources from the Anderson-Inman and Horney (2007) typology that matched the resources provided by the software used in the intervention included the following:

1. Customizable content (Presentational resource)
2. Links to resources and related documents and media (Navigational resource)
3. Alternate versions of content (Translational resource)
4. Descriptions and clarifications of content (Explanatory resource)
5. Visual representations of content (Illustrative resource)
6. Questions and testing (Evaluational resource)

Based on an analysis of the software used in this research, in relation to the two conceptual frameworks discussed (e.g., media focused vs. supportive resource), a construct of supported electronic text emerged to best describe the intervention used in this study. The research, therefore, focused on typical outcome measures associated with content area reading and related learning strategies (Higgins, Boone, & Lovitt, 2002; Readence & Moore, 1992). Indeed, the questioning format utilized in the pretest/posttest phases of this study has been emphasized as a formal and codified reading and comprehension assessment for almost 100 years (Readence & Moore, 1992).

Teaching History and Social Studies

Many of the difficulties students face in content area learning stem from mismatches between the teachers' instructional approaches and students' strategies for cognitive intake and processing of the material presented (Boone & Higgins, 2007; Chapin & Messick, 1999; Higgins, Boone, & Lovitt, 2002). These mismatches, along with disparate levels of literacy development for students within the same grade level, combine to create formidable challenges for the teaching of subject areas where student interest is already low. Mounting evidence suggests that students generally find history and social studies dull and unimportant, that they have difficulty understanding their textbooks, and that overall, they remember very little of what they learned (Ciborowski, 2005; Stetson & Williams, 2005; White, 1999). In fact, social studies and history are rated by middle school students as two of the least favorite subjects in the curriculum (Higgins, Boone, & Lovitt, 2002; Lounsbury, 1988; Shaughnessy & Haladyna, 1985), with only English receiving more negative reviews about the teaching of its content.

Technology-Supported Social Studies Learning

Educational technology and interactive multimedia play an increasingly vital role in efforts to move social studies from the rote memorization of dates and information toward a more student-centered, hands-on, authentic learning experience (Bitter & Pierson, 2005; Trinkle & Merriman, 2000). And despite movements within the discipline to promote student computer use to facilitate reflective inquiry, decision making, and problem solving (Evans, 2004; National Council for the Social Studies, 1994), social studies education for the most part continues to focus on traditional, teacher-directed, lecture-and-textbook-based approaches and activities (Diem, 2000; Friedman & Hicks, 2006; White, 1999).

The research base on the effectiveness of technology as an instructional component for teaching social studies is quite limited (Cantu, 2000). Nonetheless, there are data indicating that when integrated effectively, multimedia technology can support history and social studies learning by promoting student-centered instruction, increasing learner motivation, and extending and deepening understandings of historic and civic concepts (Molebash, 2002). Some studies have reported modest positive outcomes for several groups of students who used computer-adapted tutorial programs for the practice of social studies skills (Twyman & Tindal, 2006). In two research studies conducted a full decade apart, Higgins, Boone, and Lovitt (1996) found that hypermedia study guides resulted in positive gains for ninth grade social studies students with regard to recall, comprehension, and attitudes, while Boon et al. (2006) reported similar results in their investigation of high school students' use of technology-enhanced cognitive organizers.

METHOD

This study investigated social studies achievement as a result of utilizing a multimedia-based, American history software program to augment textbook and lecture materials for seventh-grade middle school history students. Student pretest and posttest scores on a multiple-choice assessment instrument served as the primary data source in this quasi-experimental research design.

The Early American History Software

The early American history software used in this study was an online middle school curriculum designed to help students learn the content and skills specified by state and national academic standards in a student-centered, multimedia-rich manner appealing to a wide variety of learning styles and interests (Ignite Learning, 2003). The software was a type of computer-aided instruction (CAI) that blended networked multimedia technologies for content delivery with tools to aid teachers in tracking student progress and designing individualized instruction based on the program's assessments. The program itself was Web browser-based, but self-contained in that it prevented access to the Internet and World Wide Web while the program was running. The software contained 15 units that combined multiple modalities to meet the learning objectives in each unit. The researchers had no relationship with the software company that produced the history program used in the study. A software license was purchased by the school district on a one-year trial basis, during which the district technology coordinator requested that the researchers evaluate the software. The company provided online and on-site technical assistance for the district technology coordinator, classroom teachers, and researchers for the duration of the trial period.

Research Design

This study sought to examine the correlation between multimedia software use and student outcome scores, specifically whether use of the American history software would significantly raise student achievement scores on a criterion-referenced, standards-based test. Analyses of data from 184 student test scores utilized descriptive and inferential statistical procedures to interpret the outcome-oriented test results. Pretest and posttest scores for students in control and experimental groups were compared using a two-tailed t test with unequal variance. A two-tailed t test with unequal variance was implemented because it was unclear at the time of comparison which direction mean test scores would shift and because a two-tailed t test is more sensitive to changes than a one-tailed t test.

The study aligned with the criteria for methodology, data collection, analysis, and description for scientifically based research as explicated in the No Child Left Behind Act (NCLB, 2002). Professional literature on NCLB's criteria for scientifically based research (Margolin & Buchler, 2004; NCLB, 2002; Poggi, 2003; U.S. Department of Education, 2005) guided the methodology, with Dawson's (2004) framework providing the overarching criteria for scientifically based research, which included:

1. Empirical methods are used to carry out the research, which is conducted in a systematic and consistent manner, with keen attention to detail.
2. Data collection and analysis are rigorously conducted to ensure that the data are collected, analyzed, and interpreted correctly.
3. Measurements or observational methods that provide scientifically valid and reliable measurements across many different measurement points and observations are used.
4. The studies employ experimental or quasi-experimental methodology to optimize the researchers' ability to answer the questions under investigation.
5. Enough data and description should be provided so that future researchers can attempt to replicate the findings by conducting a study using the same methods and instruments (p. 5).

Participants

Students. Subjects were seventh-grade students enrolled in public middle schools in a large urban school district in the southwestern United States. Students in eight separate sections of seventh-grade history, taught by four different teachers in three different middle schools participated in the study. The experimental group of students received treatment (i.e., use of the Ignite! program) in addition to textbook- and lecture-based instruction for all units of early American history study. The control group received textbook and lecture instruction only but did not use the Ignite! program. During both instructional conditions, the same teacher administered textbook- and lecture-based instruction in presenting the same information to both groups of students. The overall sample size was 184 pretests and posttests, obtained from an experimental group comprised of 93 students, and a control group comprised of 91 students.

Teachers. Four female teachers participated in the study. The teachers worked at three different middle schools, collectively teaching American history to a total of 637 seventh-grade students each day. Each participating teacher taught an experimental group (one full class) of students in which the American history software was used as an instructional supplement, as well as a control group (a different class) in which the software was not used. This ensured that both control and experimental group students had the same teacher, helping to reduce the chance of sampling bias. The average age of the teachers was 35 years, with an average of 9.5 years of teaching experience. Descriptive information about the participating teachers (names are pseudonyms) is shown in Table 1. Table 2 provides information about the schools participating in the study (names are pseudonyms).

Setting. The study was conducted in a large, rapidly growing school district in the Southwest, in which approximately 40.4% of seventh grade students qualified to receive free or reduced lunch, 14.92% of seventh graders were non-English proficient or had limited English proficiency, and 11.1% received special education services under an individualized education program (IEP). The district's student population was approximately 14% African American, 33.4% Hispanic, and 43.9% Caucasian. However, two of the three schools included

Table 1: Descriptive Data for Participating Teachers (all are pseudonyms)

Teacher Name	Middle School	Age	Years Teaching	Highest Degree
Romero	Samuels MS	30	5	B.A.
Gage	Hawthorne MS	55	26	M.A.
Smith	Hawthorne MS	31	7	M.A.
Brown	Jackson MS	24	0	B.A.

Table 2: Middle School Student Demographic Info (all are pseudonyms)

Teacher	Middle School	% of LEP Students	% of IEP Students	% Eligible for Free or Reduced Lunch
Romero	Samuels	22.5	11.2	61.8
Gage	Hawthorne	17.5	12.0	48.6
Smith	Hawthorne	17.5	12.0	48.6
Brown	Jackson	6.7	12.2	31.7

in this study had a higher than average rate of seventh graders eligible to receive free or reduced lunch: 50.7% and 61.8%, with minority populations of 61.6% and 56.2% respectively. The schools that participated in the study were geographically distributed throughout the district, with every effort made to select equivalent teachers and students for the treatment and control groups who were representative of the district's typical student population.

Eligibility for school participation. In selecting the sites for the study, participation was limited to schools with adequate technology infrastructure, computer facilities, and interested teachers. This decision was made after reviewing findings from a pilot study (Kingsley, 2003) to determine potential difficulties and problems associated with implementing the program in schools on a larger scale. Results from the pilot study indicated that schools lacking high-speed, high-capacity server and networking capabilities were frequently plagued with server and work station crashes, software freezes, very slow response time, and/or inability for students to run all of the media segments contained in the program. As random selection of participating schools was not possible, the decision was made to follow Stake's (2000) heuristic that in some cases, the opportunity to learn from a site should take priority over concern for its typicality of an entire population.

Teacher training. At the start of the school year, all participating teachers attended mandatory introductory training that provided an overview of the American history software and familiarized the teachers with available content and media options. In the training session teachers learned how to construct assignments, select assessments, create new sections for classes using the program, create student logins and passwords, and how to locate and use the multimedia options.

At the start of the school year, the teachers conducted an orientation session to show students in the experimental groups how to log in to the program and set their passwords and demonstrated the program's content options and navigational aids. At that time, the students were given teacher-facilitated, hands-on time to familiarize themselves with login procedures and with the program's interface, functionality, and media choices prior to commencement of actual instruction. Additionally, all students in the experimental and control groups received an overview of the history textbook to be used in the course, an outline of the seventh grade history curriculum, and a syllabus for the entire history course.

Data Collection

Procedure. Prior to using the American history software, each participating teacher designated one of her classes to be a treatment group, and another, similar class as a control group of students. In all cases, the treatment and control classes were inclusive, homogenous, general education seventh-grade history classes to which students had been assigned independently of teacher or researcher oversight or influence. The overall sample size was 184 pretests and posttests, obtained from an experimental group comprised of 93 students, and a control group comprised of 91 students. Every effort was made to select groups that would be as similar as possible; however, the use of intact classrooms often results in Non-Equivalent Group Design (NEGD). NEGD, a common feature of social and educational research, is particularly susceptible to internal validity and selection maturation threats, which require further statistical analyses (Hancock, Knezek, & Christensen, 2007). Two additional reliability tests for internal consistency (e.g., *Split-Half Correlation* and *Cronbach's Alpha*) were conducted to support and validate the findings of this study.

Instructional conditions. The experimental group of students received treatment (i.e., use of the American history software) in addition to textbook and lecture-based instruction for all units of study. The control group received textbook and lecture instruction but did not use the American history software. During both instructional conditions, the same teacher administered textbook and lecture-based instruction in presenting the same information to both groups of students.

With the treatment group students, teachers reserved a minimum of 20% of the instructional time, or approximately one day per week, for use of the American history software. Regular textbook instruction consisted of using either *The American Journey* (Appleby, Brinkley & McPherson, 2003), or *The American Nation* (Davidson, Castillo, & Stoff, 2002). Both district-approved books were similar in content, scope, and sequence of information. The textbooks included graphic organizers and other visual aids, such as timelines, photographs and illustrations, and political maps, as well as vocabulary lists, chapter outlines, and chapter summaries. Teachers supplemented book-based instruction with online and offline auxiliary activities provided by the textbook publishers, as well as with their own materials, worksheets, and selected Web sites. No other instructional software programs were used for history instruction during the study.

As specified by district policy, students had a copy of their history textbook at home, and each classroom had another set for student use at school. Students were unable to access the American history software from home. In both the experimental and control groups, the curriculum requirements were identical and were based on the state history standards scope and sequence.

The procedure for both instructional groups from pretest to posttest conditions lasted approximately 7 months. Classes consisted of 50-minute block periods encompassing daily review, learning objectives, presentation of new information, and in some cases independent practice. On days the software was used by students in the experimental groups, class sessions consisted of students navigating through the assigned lesson in any order and at their own pace, provided that they viewed all of the media contained in the assigned module. After viewing the media pieces for the assignment, students completed a Topic Review: a six-item multiple-choice assessment built into each lesson. Scores from the Topic Reviews were not used for the current study; rather, they served as an instructional focal point for students while they used the program. In each 50-minute class period where the software was used, students were able to complete one full lesson and its accompanying Topic Review. Upon completion of the early American history portion of the history course, student participants were given the 50-item posttest to measure their knowledge and recall of major concepts related to the period of American history from 1492 to 1877 (i.e., the period of Reconstruction).

Instrument. Because the quizzes and topic reviews contained in the software were closely tied to that specific content, they were not used as a measure of achievement for the study. Rather, an independent, criterion-referenced pretest was administered to all participating students at the onset of the seventh-grade school year. Material on the pretest consisted of knowledge required to master the seventh-grade history curriculum as outlined by state standards. The full pretest instrument consisted of 50 multiple-choice questions based on the state's scope and sequence history standards. An identical posttest was administered at the conclusion of the 7-month instructional period.

The multiple-choice pretest-posttest instrument included questions drawn from a test bank of 4,500 questions accompanying *The American Journey* (Appleby, Brinkley & McPherson, 2003) history textbook, as well as questions contributed by several history teachers in the participating middle schools. Because multiple-choice tests tend to focus on basic facts and are not always good measures of higher level cognitive processing (Becker, 1992), some multiple-choice questions on the pretest and posttest were adjusted slightly to facilitate problem-solving, decision making, and/or higher order thinking skills related to the concepts covered in the history knowledge being tested. The pretest-posttest instrument was compiled by three researchers with experience in designing and conducting education research and evaluation who were familiar with this research project. Reliability checks on the instrument were conducted independently by the test designers, and discrepancies were discussed and assessed to obtain 100% agreement. In addition, due to the identification of NEGD in this study, two additional post hoc reliability tests for measuring internal consistency were performed on this instrument. The first test, a Split-Half Correlation, randomly divides all the instrument items into two sets and measures the correlation between the total scores for each randomly divided half of the questions, which resulted in a high internal validity score (α=.99) that greatly exceeds the acceptable standard, α >0.90. The second, Cronbach's Alpha, which tends to be a more stringent test, revealed high reliability for the control (α=.98) and experimental groups (α=.99), as well as the comparison of all groups using the pre-test analysis (α=.98). The instrument's concurrent validity with questions from the test bank of questions drawn from the district-approved textbook *The American Journey* was checked and obtained a high validity coefficient (.87). The instrument was examined and approved by the district technology coordinator, the district social studies coordinator, and two of the most experienced participating history teachers for content validity to ensure high correlation with the scope and sequence of American history content as specified in the state curriculum standards. It was then pilot tested with a small sample of doctoral students before the study began.

Use of the American History Software

Throughout this study, qualitative data were collected to document the degree of fidelity with which the history software was used by all participating teachers and students. Weekly classroom observations were augmented by conversations and informal interviews with the teachers throughout the 7-month investigation period. Transcripts from observations and teacher interviews revealed that each of the participating teachers used the program for the equivalent of one class period per week throughout the period of investigation.

Students used the history software in a computer lab where they had access to their own computers and were free to work through the assigned modules at their own pace. On occasion, the history classes were usurped by another group of students who needed to use the computer lab. In these cases the teachers requested an extra day the following week in order to recoup the missed lab time. On a few occasions when the computer lab was occupied on their assigned lab day, the teachers used the American history software as a teacher-directed, whole-class instructional tool. In these rare cases, the teacher led a discussion of the materials and showed the media pieces to the students on a large projection screen in the class, while students took notes or completed an outline of the material covered. However, the vast majority of time spent using the American history software was weekly time in the computer lab with students engaged one-to-one with the program on a computer, setting their own pace and using the media pieces in any order they preferred.

Data Analysis

Descriptive and inferential statistics, including mean, standard deviation, and two-tailed *t* tests, were used on the pretest and posttest scores for students in the experimental and control groups. Statistical analyses were conducted using the Statistical Product and Service Solutions (SPSS) software to determine the significance of variables related to the research question. According to Valdez (2004), educational researchers, especially those who have conducted metaanalyses, agree that when used appropriately, technology can improve education in the effect-size range of between 0.30 and 0.40 (Kulik, 2002). Cohen (1977) classified effect sizes of around 0.2 as small, around 0.5 as moderate, and around 0.8 as large (p. 1). In order to obtain a power rating of .80 with an effect size of .50 (moderate effect), there needed to be at least 50 students in each of the control and treatment groups, assuming use of a two-tailed test with an alpha of .025 (Gay & Airasian, 2000). With sample numbers of more than 50 for each group, it was possible to measure lesser effects. Two-tailed *t* tests were used because it was unknown whether effects from using the American history software would be positive or negative.

RESULTS

This investigation examined whether there was a statistically significant difference between pretest and posttest achievement scores for students who used the American history software compared to students who did not use the program. Using the computer software program Statistical Product and Service Solutions (SPSS), descriptive and inferential statistics were compiled from the pretest and posttest scores of students in control and experimental groups. The mean scores of control and experimental groups on pretests and posttests were calculated, then compared using a two-tailed *t* test with unequal variance.

Pretest. For students in the pretest control group (n= 91), the average number of correct answers was 33.60 out of 50 total questions with a standard deviation of 5.30, while the average number of correct answers for all students in the pretest experimental group (n=93) was 30.95 out of 50 total questions, with a standard deviation of 6.12. In other words, students in the control group had a 67.2% pretest average for correct answers, while students in the experimental group had a pretest average of 61.9% for correct answers.

Posttest. At the end of the instructional period being studied, the average number of correct answers for students in the posttest control group (n=91) was 36.66 out of a total of 50 questions with a standard deviation of 5.58, the equivalent of 73.32% correct, while the average number of correct answers for students in the posttest experimental group (n=93) was 37.04 of 50 total questions with a standard deviation of 5.51, the equivalent of 74.07% correct.

The mean posttest scores indicated that students who used American history software, as well as those who did not use it, both increased their test scores from pretest to posttest conditions. These data describe the classic "selection-maturation threat" potential of any NEGD project, in which improvement was observed in both the "non-equal" experimental and control groups. However, examination of the percentage increase between pretest control and pretest experimental groups to posttest control and posttest experimental group revealed that students in the control group increased their mean test scores an average of 6.1%, while students in the experimental group increased their mean test scores an average of 12.2%, or approximately twice as much. This difference in mean test scores was statistically significant. Moreover, although the reliability measures (i.e., Split-Half, Cronbach's Alpha) demonstrated reliability for all groups for the pre-test—post-test reliability was significantly reduced (α=0.48), providing further evidence that suggests the experimental treatment was sufficient to significantly alter the outcomes measured by this instrument.

The significance level associated with the difference in test score results between the control and experimental groups was less than 0.01%. The question less confidently answered was the likelihood of attribution of the treatment to the 12.2% mean test score increase for students in the experimental group versus the 6.1% mean test score increase for students in the control group.

DISCUSSION

This section addresses the research question: Was there a significant difference between pretest and posttest achievement scores for students who used the American history software as compared to students who did not use the program? Results indicated statistically significant positive effects on overall achievement scores for students who used the American history software. Mean test scores for students who used the software improved by 12.2% and an average of 6.09 more correct answers from pretest to posttest, while mean scores for control group students improved by 6.1%, an average of 3.06 more correct answers from pretest to posttest. The significance level in a statistical study is the risk associated with not being 100% confident that what was observed in an experiment or quasi-experiment was due to the treatment or what was being tested. In this case the treatment was student usage of the American history software program. Since the impact of all other potential factors on the differences observed between outcomes of treatment and control groups cannot be eliminated confidently, some level of probability (i.e., the p value) is assigned and reported. On a two-tailed t test of unequal variance, a very high level of significance was found, $p=0.0000000337623$, where p represents the probability that the increase in mean test scores was attributable to something other than use of the American history software. The data suggest that the difference in outcome scores for students in the experimental group was likely due to their use of the American history software.

In addition to reporting outcomes and probability levels for errors, studies conducted in consonance with the No Child Left Behind (2002) definition for scientifically based research must also report the effect size and statistical power of a study. Statistical power is related to the variance: the smaller the variation relative to each group (e.g., between the experimental and control groups), the larger a sample size must be in order to obtain a high power rating. The power of a statistical hypothesis test measures the test's ability to reject the null hypothesis when it is actually false—that is, to make a correct decision (Hinkle, Wiersma, & Jurs, 1998). Obviously, the higher the power rating, the more reliable the statistical test. The maximum power a test can have is 1, and the minimum is 0. Ideally, researchers would strive to have a high power or a number close to 1. In this study, for the control group of students the power was 0.965, and for the experimental group the power was 1.00. In other words, there was a 96.5% statistical likelihood that the two-tailed t test was able to detect the effects for the control group of students, and a 100% chance that it was able to detect the effects for the experimental group. Consequently, it can be asserted that the results of the two-tailed t test on both the control and experimental groups yielded valid, reliable results. Overall, results of this quasi-experiment suggest a strong link between use of a technology-enhanced intervention and higher outcome achievement scores for this group of middle school learners.

Limitations of the Study

Because participants in this study were not randomly selected but were instead part of a cohort, the generalizability of the results to similar student populations is considered lower than if the sampling process had been completely random. The non-experimental group design (NEGD) of this study precludes the drawing of causal inferences. Generalizability may also be limited by the fact that all of the participating teachers were female.

Implications and Future Research

This study was designed and conducted in consonance with criteria specified in the No Child Left Behind Act's (2002) definition of scientifically based research. With instructional technology playing an increasingly central role in all academic areas, more research, and more effective approaches are needed to document student achievement related to computer-based training and educational programs (Bull et al., 2005; U.S. Department of Education, 2005). This study adds to the body of scientifically based research literature on student achievement directly linked to the use of educational software. Bull et al. describe the compelling need for this sort of research by stating "[t]o date there have been no documented systemic increases in student achievement and learning directly attributable to technological innovation" (p. 218). They add, "[t]here is no area in which well-conceived and effectively implemented research could be of greater value than in the area of [educational] technological innovation" (p. 218). The current study responds to calls for accountability from scholars, policymakers, and educators at all levels for rigorous evidence indicating whether technology investments can truly support student learning (Jones et al., 2004–2005) in educational settings. Furthermore, this study adds to the very limited body of research on the effectiveness of technology as a component for teaching social studies (Cantu, 2000; Diem, 2000).

Results of this study suggest several potential directions for further exploration. One possible avenue for investigation addresses concerns that a major shortcoming of scholarly literature on the efficacy of technology in education is that the research varies tremendously in methodology, sampling, and focus. Researchers (Kirkpatrick & Cuban, 1998; Waxman, Lin & Michko, 2003) have documented studies with large variations in sampling, such as differences in student grade levels, socioeconomic classes, and aptitudes. The current study employed a quasi-experimental methodology that implemented all NCLB criteria for scientifically based research, including a disparate sampling of teachers and students from the school district, and a strong emphasis on using the instructional intervention with fidelity. An experimental research design employing in-depth qualitative and quantitative data analyses would provide greater insight into the causal factors surrounding the differences in student achievement, while also providing information related to the processes teachers use to integrate new technology into their existing curriculum.

In the current study, the data suggest that use of a software program affected student achievement scores on a standards-based, multiple-choice test; however, many questions about the effects of educational software on student learning remain unanswered. Another possible direction for further research would be to investigate gains in student achievement if the program were to be used with students for more than the 20% of instructional time implemented in this study. For instance, students who did not use the software showed an average mean test score increase of around 6%, while those who used the program had mean test score increases of about 12%, or twice as much. It would be interesting to investigate what might happen if software usage were increased to 25%, or even to 50% of instructional time. Would students continue to show gains on standardized assessments, or would a point of diminishing returns be reached?

Final directions for further investigation include exploring whether test score increases attributed to use of the history software would be enough to truly make a difference in whether students pass their seventh grade history courses (i.e., practical significance vs. statistical significance) and whether knowledge gained from use of the software transfers to more complex learning tasks outside the context of a standardized written examination.

Contributors

Karla V. Kingsley is an assistant professor of instructional technology and educational computing at the University of New Mexico. Her current research focuses on technology-mediated interactive learning, social semiotics, digital game-based learning (DGBL), and critical pedagogy. (Address: Karla V. Kingsley, PhD, University of New Mexico Department of Teacher Education, MSC05-3040, Albuquerque, NM 87131-1231; E-mail: karlak@unm.edu).

Randall Boone is a professor of educational computing and technology at the University of Nevada Las Vegas. His current research focuses on instructional and assistive technologies for persons with both learning and sensory disabilities. (E-mail: randall.boone@unlv.edu).

References

Alessi, S. M., & Trollip, S. R. (2001). *Multimedia for learning: Methods and development* (3rd ed.). Boston, MA: Allyn & Bacon.

Anderson-Inman, L., & Horney, M. (1998). Transforming text for at-risk readers. In D. Reinking, L. D. Labbo, M. C. McKenna, & R. D. Kieffer (Eds.), *Handbook of literacy and technology: Transformations in a post-typographic world* (pp. 15–43). Mahwah, NJ: Erlbaum.

Anderson-Inman, L., & Horney, M. (2007). Supported etext: Assistive technology through text transformations. *Reading Research Quarterly*, *42*(1), 153–160.

Appleby, J., Brinkley, A., & McPherson, J. M. (2003). *The American journey*. New York: Glencoe McGraw-Hill.

Bagui, S. (1998). Reasons for increased learning using multimedia. *Journal of Educational Multimedia and Hypermedia, 7*(1), 3–18.
Becker, H. J. (1992). Integrated learning systems and their alternatives: Problems and cautions. *Educational Technology, 32*, 51–57.
Becker, H. J. (2000). Findings from the Teaching, Learning, and Computing survey: Is Larry Cuban right? *Education Policy Analysis Archives 8*(51). Available at http://epaa.asu.edu
Bitter, G. G., & Pierson, M. E. (2005). *Using Technology in the Classroom* (6th ed.). Boston, MA: Pearson Education.
Boon, R. T., Burke, M., Fore, C., Spencer, V. (2006). The impact of cognitive organizers and technology-based practices on student success in secondary social studies classrooms. *Journal of Special Education Technology, 21*(1), 5–15.
Boone, R., & Higgins, K. (2005). Designing digital materials for students with disabilities. In D. Edyburn, K. Higgins, & R. Boone (Eds.), *Handbook of special education technology research and practice* (pp. 481–492). Whitefish Bay, WI: Knowledge by Design, Inc.
Boone, R., & Higgins, K. (2007). The role of instructional design in assistive technology research and development. *Reading Research Quarterly*, *42*(1), 135–140.
Brouwer, N., Muller, G., & Rietdijk, H. (2007). Educational designing with MicroWorlds. *Journal of Technology and Teacher Education, 15*(4), 439–462.
Brünken, R., Plass, J., & Leutner, D. (2003). Direct measurement of cognitive load. *Educational Psychologist, 38*(1), 53–61.
Bull, G., Knezek, G., Roblyer, M. D., Schrum, L., & Thompson, A. (2005). A proactive approach to a research agenda for educational technology. *Journal of Research on Technology in Education, 37*, 217–220.
Campbell, D. T. (1969). Reforms as experiments. *American Psychologist, 24*, 409–429.
Cantu, D. A. (2000). Technology integration in pre-service history teacher education. *Journal of the Association for History and Computing, 3*(2). Retrieved on May 2, 2007 from http://mcel.pacificu.edu/JAHC/JAHCIII2/K12/cantu.htm
Chapin, J. R., & Messick, R. G. (1999). *Elementary Social Studies* (4th ed.). New York: Longman.
Ciborowski, J. (2005). Textbooks and the students who can't read them. In G. Moss (Ed.), *Critical reading in the content areas* (pp. 206–216). Dubuque, IA: McGraw-Hill/Dushkin.
Clark, R. C., & Mayer, R. (Eds.). (2003). *E-learning and the science of instruction: Proven guidelines for consumers and designers of multimedia learning*. San Francisco: Pfeiffer.
Cohen, J. (1977). *Statistical Power Analysis for the Behavioral Sciences* (2nd ed.). New York: Academic Press.
Cook, J. (2001). Bridging the gap between empirical data on open-ended tutorial interactions and computational models. *International Journal of Artificial Intelligence in Education, 12*(1), 85–99.
Crosier, J., Cobb, S., & Wilson, J. (2002). Key lessons for the design and integration of virtual environments in secondary science. *Computers & Education, 38*, 77–94.

Davidson, J. W., Castillo, P., & Stoff, M. B. (2002). *The American Nation*. Upper Saddle River, NJ: Prentice Hall.

Dawson, M. (2004). NCREL Quick Key No. 7: A foundation for understanding and evaluating scientifically based research. *Educational Technology News, 4*(1), 5. Naperville, IL: North Central Regional Educational Laboratory. Retrieved January 10, 2006, from http://www.ncrel.org/csri/tools/qkey7/index.html

Diem, R. A. (2000). Can it make a difference? Technology and the social studies. *Theory & Research in Social Education, 28*(4), 493–501.

Evans, R. W. (2004). *The social studies wars: What should we teach the children?* New York: Teachers College Press.

Friedman, A. M., & Hicks, D. (2006). Guest editorial: The state of the field: Technology, social studies, and teacher education. *Contemporary Issues in Technology in Teacher Education, 6*, 246–258.

Gay, L. R., & Airasian, P. (2000). *Educational research: Competencies for analysis and application* (6th ed.). Upper Saddle River, NJ: Prentice Hall.

Gibbs, W., Graves, P., & Bernas, R. (2001). Educational guidelines for multimedia courseware. *Journal of Research on Technology and Education, 34*(1), 2–18.

Greenlee-Moore, M. E., & Smith, L. L. (1996). Interactive computer software: The effects on young children's reading achievement. *Reading Psychology, 17*(1), 43–64.

Hancock, R., Knezek, G., & Christensen, R. (2007). Cross-validating measures of technology integration: A first step toward examining potential relationships between technology integration and student achievement. *Journal of Computing in Teacher Education, 24*(1), 15–21.

Higgins, K., & Boone, R. (2001). Adapting instruction for children with disabilities. In L. W. Searfoss & J. E. Readence (Eds.), *Helping children learn to read: Creating a classroom literacy environment (*4th ed. pp. 330–365). Boston: Allyn and Bacon.

Higgins, K., Boone, R., & Lovitt, T. (1996). Hypertext support for remedial students and students with learning disabilities. *Journal of Learning Disabilities, 29*(4), 402–412.

Higgins, K., Boone, R., & Lovitt, T. C. (2002). Adapting challenging textbooks to improve content area learning. In G. Stoner, M. R. Shinn, & H. Walker (Eds.), *Interventions for achievement and behavior problems* (2nd ed., pp. 755–790). Silver Spring, MD: National Association for School Psychologists.

Hinkle, D. E., Wiersma, W., & Jurs, S. G. (1998). *Applied Statistics for the Behavioral Sciences* (4th ed.). Boston, MA: Houghton Mifflin Company.

Horney, M., & Anderson-Inman, L. (1999). Supported text in electronic reading environments. *Reading and Writing Quarterly*, *15*(2), 127–168.

Ignite! Early American History. (2003). [Computer software]. Austin, TX: Ignite! Learning.

Ignite! Learning. (2003). *Teaching students in the ways they learn best: The Ignite!method of instructional design.* Retrieved January 22, 2005, from http://www.Ignite!learning.com/methodology.shtml

Jones, J. D., Staats, W. D., Bowling, N., Bickel, R. D., Cunningham, M. L., & Cadle, C., (2004–2005). An evaluation of the Merit Reading Software Program in the Calhoun County (WV) Middle/High School. *Journal of Research on Technology in Education, 37*(2), 177–195.

Kingsley, K. V. (2003). *Evaluating the viability of the Ignite! Early America History program in middle schools: A pilot study*. Unpublished manuscript, University of Nevada Las Vegas.

Kirkpatrick, H., & Cuban, L. (1998). Computers make kids smarter—right? *Technos Quarterly, 7*(2). Retrieved June 20, 2005, from http://www.technos.net/tq_07/2cuban.htm

Kulik, J. A. (2002). *School mathematics and science programs benefit from instructional technology* (InfoBrief). Washington, DC: National Science Foundation. Retrieved June 22, 2007, from http://www.nsf.gov/sbe/srs/infbrief/snf03301/start.htm

Labbo, L. (2002). Computers, kids, and comprehension: Instructional practices that make a difference. In C. C. Block, L. B. Gambrell, & M. Pressley, (Eds.), *Improving comprehension instruction: Rethinking research, theory, and classroom practice* (pp. 275–289). San Francisco, CA: Jossey-Bass.

Lockard, J., & Abrams, P. D. (2004). *Computers for twenty-first century educators* (6th ed.). Boston, MA: Allyn & Bacon.

Lounsbury, J. (1988). Middle-level social studies: Points to ponder. *Social Education, 52*(2), 116–118.

Margolin, J., & Buchler, B. (2004). *Critical Issue: Using scientifically based research to guide educational decisions.* Naperville, IL: North Central Regional Educational Laboratory. Retrieved on January 10, 2005 from http://www.ncrel.org/sdrs/areas/issues/envrnmnt/go/go900.htm

Mayer, R. E. (2003). The promise of multimedia learning: Using the same instructional design methods across different media. *Learning and Instruction, 13,* 125–139.

Mayer, R. E. (2005). Introduction to multimedia learning. In R. E. Mayer (Ed.), *The Cambridge handbook of multimedia learning* (pp. 1–18). New York, N.Y: Cambridge University Press.

Mills, R. J. (2001). Analyzing instructional software using a computer-tracking system. *Information Technology, Learning, and Performance Journal, 19*(1), 21–30.

Molebash, P. E. (2002). *Constructivism meets technology integration: The CUFA* technology guidelines in an elementary social studies methods course. *Theory and Research on Social Education, 30*(3), 429–455.

Moreno, R. (2006). Learning with high tech and multimedia environments. *Current Directions, 15*, 63–67.

Moreno, R., & Valdez, A. (2005). Cognitive load and learning effects of having students organize pictures and words in multimedia environments: The role of student interactivity and feedback. *Educational Technology Research & Development, 53*(3), 35–45.

National Council for the Social Studies. (1994). *Expectations of excellence: Curriculum Standards for Social Studies*. Washington, DC: Author.

National Research Council. (2002). *Scientific research in education.* Washington, DC: National Academy Press.
No Child Left Behind Act of 2001, Pub. L. No. 107–110, 115 Stat. 1425. (2002). Retrieved December 1, 2007, from http://www.ed.gov/policy/elsec/leg/esea02/index.html
Poggi, S. (2003). *Wake-up call: Facing the challenge to use scientifically based research in schools.* Naperville, IL: North Central Regional Educational Laboratory. Retrieved January 10, 2005, from http://www.ncrel.org/info/nlp/lpsp03/index.html
Readence, J. E., & Moore, D. W. (1992). Why questions? A historical perspective on standardized reading comprehension tests. In E. K. Dishner, T. W. Bean, J. E. Readence, & D. W. Moore (Eds.), *Reading in the content areas* (3rd ed., pp. 390–398). Dubuque, IA: Kendall/Hunt Publishing Company.
Roblyer, M. D. (1999). Our multimedia future: Recent research on multimedia's impact on education. *Learning & Leading with Technology, 26*(6), 51–54.
Shaughnessy, J. M., & Haladyna, T. M. (1985). Research on student attitudes toward social studies. *Social Education, 49,* 692–695.
Stake, R. (2000). Case studies. In N. K. Denzin & Y. S. Lincoln (Eds.) *Handbook of qualitative research* (2nd ed., pp. 435–454). Thousand Oaks, CA: Sage.
Stetson, E., & Williams, R. (2005). Learning from social studies textbooks: Why some students succeed and others fail. In G. Moss (Ed.), *Critical reading in the content areas* (pp. 135–142). Dubuque, IA: McGraw-Hill/Dushkin.
Thompson, A. (2007). Scientifically based research: Establishing a research agenda for the technology and teacher education community. *Journal of Research on Technology in Education, 37*(4), 331–338.
Trinkle, D. A., & Merriman, S. A. (2000). *The history highway 2000: A guide to Internet resources* (2nd ed.). Armonk, NY: M.E. Sharpe.
Twyman, T., & Tindal, G. (2006). Using a computer-adapted, conceptually based history text to increase comprehension and problem-solving skills of students with disabilities. *Journal of Special Education Technology, 21*(2), 5–16.
U.S. Department of Education. (2005). *Toward a new golden age in American education: How the Internet, the law and today's students are revolutionizing expectations.* Washington, DC: Office of Educational Technology.
Valdez, G. (2004). *Critical issue: Technology leadership: Enhancing positive educational change.* Naperville, IL: North Central Regional Educational Laboratory. Retrieved January 3, 2006, from http://www.ncrel.org/sdrs/areas/issues/educatrs/leadrshp/le700.htm
van Merriënboer, J. G., & Sweller, J. (2005). Cognitive load theory and complex learning systems: Recent developments and future directions. *Educational Psychology Review*, *17*(2), 147–177.
Waxman, H. C., Lin, M., & Michko, G. M. (2003). *A meta-analysis of the effectiveness of teaching and learning with technology on student outcomes*. Naperville, IL: Learning Point Associates.
White, C. W. (1999). *Transforming Social Studies Education: A Critical Perspective*. Springfield, IL: Charles C. Thomas.

Williams, D. W, Boone, R., & Kingsley, K. V. (2004). Teacher beliefs about educational software: A Delphi study. *Journal of Research on Technology in Education, 36*(3), 213–230.