International Journal of Advanced Education and Research
ISSN: 2455-5746, Impact Factor: RJIF 5.34
www.alleducationjournal.com
Volume 1; Issue 12; December 2016; Page No. 30-36

Using artificial intelligence to enhance the effectiveness of multimedia-based instruction

[1] John Leddo, [2] Darika Bisht, [3] Esha Narla, [4] Rishi Saranu, [5] Max Titov

[1] MyEdMaster and Education Online, Inc., Herndon, Virginia 20171, USA
[2, 3, 4, 5] MyEdMaster LLC, Herndon, Virginia 20171, USA

Abstract

The Internet has given rise to a plethora of multimedia educational resources that are available to students. Two of the most popular instructional formats are video-based instruction and educational television. These are attractive since they provide engaging education and reach wide audiences. However, a potential weakness is that the type of personalized instruction a teacher provides is typically lost. The present paper investigates whether combining these formats with artificial intelligence (AI) technology can enhance student performance. Two experiments were conducted, one that added AI to Internet-delivered educational TV and one that added AI to Internet-delivered VBI. In both cases, student performance was dramatically improved. The results suggest that AI can improve educational performance when added to the multimedia-based instruction.

Keywords: internet, multimedia educational, artificial intelligence (AI)

Introduction

The past two decades have seen an enormous increase in the use of technology to supplement and even replace traditional classroom instruction. Nowhere is this more prevalent than with the Internet where there are now countless sites, many of them free, that offer supplemental educational resources for students. Technology-based education, often referred to as e-learning, has received a lot of attention in the research community. E-learning has increased in popularity over the years because it provides benefits such as: provides time and location flexibility; results in cost and time savings for educational institutions; fosters self-directed and self-paced learning by enabling learner-centered activities; creates a collaborative learning environment by linking each learner with physically dispersed experts and peers; allows unlimited access to electronic learning material; and allows knowledge to be updated and maintained in a more timely and efficient manner (Baloian *et al*., 2000; Kumar *et al*., 2001; Piccoli *et al*., 2001) [3, 11, 14].

Two of the most popular forms of supplemental instruction resources are video-based instruction (VBI) and educational television (TV). VBI is generally well-received by students. Accordingly, much research has been devoted to studying its effectiveness. Adding videos was shown to improve performance in third and eighth graders (Boster *et al*., 2006). Videos have been shown to improve educational achievement compared to other formats such as text (Khan *et al*., 2010), lecture (Siegel *et al*., 1997) [17], and simulation (Morgan *et al*., 2002) [13].

One of the reasons why videos may enhance learning is that videos tend to be engaging, which fosters student learning (Roblyer and Edwards, 2001) [15]. Videos lead to higher acceptance and satisfaction among students (Donkor, 2011) [8]. Such engagement and acceptance leads to greater attention to the material being taught and deeper cognitive processing of the material, thus enhancing learning (Balslev *et al*, 2005) [4].

The effectiveness of videos as part of instruction is enhanced when the videos are interactive as demonstrated by Zhang *et al*. (2006) [22]. Zhang and his colleagues investigated four conditions: a classroom environment, an e-learning environment with no video, non-interactive videos that students had no control over, and interactive videos where users could control what they saw and when they saw it. Results showed that both learning satisfaction and performance was greatest in the interactive video condition.

As with VBI, Internet-based TV is gaining increased interest from both the research community (Carey, 2003; Claros and Cobos, 2012; Van Tassel, 2001) [6, 7, 21] and the commercial marketplace. While websites that provide multimedia instruction are widely popular and freely available, the very prevalence of such resources gives rise to the basic question of how effective are they and is there a way to make them even better?

One possibility, which is the focus of the present paper, is through the incorporation of artificial intelligence (AI) into the multimedia-based instruction paradigm. The combination of multimedia-based instruction and AI seems promising for an online e-learning resource aimed at a wide audience. The multimedia offers the chance to teach students in an engaging and effective manner and the AI technology offers a means emulate the expertise of a human teacher to guide the learning process. For example, instead of simply requiring students to enter an answer and check to see if the answer is correct, students can be given the opportunity to enter their work and be shown what their mistakes are and how to correct them. We hypothesize that this would lead to greater learning performance compared to multimedia-based instruction alone.

In order to test this hypothesis, we compared systems that embed AI into both educational TV and VBI platforms against those that did not. Mathematics was chosen since there is evidence in the literature that AI-based technology can lead to enhanced mathematics performance (Steenbergen-Hu and Cooper, 2013) [19]. If the hypothesis is correct, students using the AI-based technologies will show higher performance than those who do not.

Adding AI to Multimedia Technology

There are essentially two major components of an AI-based multimedia system: the multimedia component and the AI component. There are variations on how each component can be created. Since the focus of the present paper is not on innovations in multimedia per se, we will not discuss the general process of creating multimedia instruction, but rather how AI can be added to that instruction. There are two broad considerations in how to do this. The first is what AI model will be used. The second is how to integrate the AI into the multimedia technology. These are discussed in turn.

AI-based Instruction

Researchers in the field of AI have focused in making technology-based instruction act more like a human teacher by encoding the deep knowledge regarding subject matter, teaching methods, and assessment abilities into the technology itself. This field is often referred to as intelligent tutoring systems (Graesser *et al*., 2012; Sottilare *et al*., 2013) [18, 18]. Many intelligent tutoring systems (ITSs) are modeled after John Anderson's ACT-R theory (1990) [1], which focuses on how people learn to proceduralize the knowledge they are taught so that they can apply that knowledge to practical problem solving. Accordingly, ITSs that are modeled after ACT-R start by giving students lessons that describe the concepts they need to learn and follow this instruction by having the students engage in step-by-step problem solving. A key component of this approach is ensuring that students correctly execute the procedures that they are taught. To accomplish this, the student's work is recorded and compared to a protocol of correct solutions. When students deviate from the correct approach, they are given feedback on what they need to do to correct their mistakes (cf., Graesser *et al*., 2012) [9].

The AI component of the technology used in the present studies is based on John Anderson's ACT-R framework (Anderson, 1990) [1] that has formed the basis of numerous AI-based instructional systems. The core of ACT-R is a production rule system where sequential procedures are stored based on the antecedent conditions that trigger them. The system then matches the student's input to the step that is listed in the production rule sequence. A match is considered to be a correct step and a mismatch is considered to be an incorrect step. ACT-R allows for more than one pathway to a solution, which is beneficial to the topics taught in the present studies since there is generally more than one way to solve a problem. Typically, people who build AI-based systems for education that are modeled after ACT-R enumerate each problem solving path that is possible for solving the problem. This is done for each specific problem that the system will deliver (cf., Aleven *et al*., 2006) [2]. This becomes particularly cumbersome if the software will ultimately deliver many problems (as would any large scale educational system) or if the system is intended to be flexible enough to allow students to enter their own homework or test-study guide problems (as we intend to allow in future versions of our system).

Therefore, in order to create a more flexible system that can support any problem within a problem class, the technology used in the present experiments operates on generalized problem types where the numbers used in the underlying production rule model the AI engine uses are parameterized rather than instantiated. For example, a typical ACT-R system might model a simple solution path for adding (2 + 3i) + (3 + 4i) as

Step 1: (2 + 3i) + (3 + 4i)
Step 2: (2 + 3) + (3i + 4i)
Step 3: 5 + 7i.

This would require a separate model for every possible problem that the system would deliver to a student. Parameterizing each variable results in a system that requires only one knowledge model per problem type plus the particular variable values for each problem. Therefore, the new solution path for the same problem looks like

Step 1: (a + bi) + (c + di)
Step 2: (a + c) + (bi + di)
Step 3: evl(a+c) + evl(b+d)i. (evl means to evaluate the sum of a+c)

Problem 1: a=2, b=3, c=3, d=4, and so on for each problem to be used.

This method means that the system can generate unlimited problems to present to the students and the AI technology can respond to them since its representation of the problem is generic rather than hardcoded. For each possible step, there are multiple pathways that are permissible and the algorithm is supplemented with mathematical expression evaluators that recognize equivalent inputs (e.g., a+bi and bi+a are mathematically equivalent).

For each step in the process, the possible errors a student could make are enumerated. For each error, there is associated text that describes the error and the way to correct it. Similarly, three hints, each progressively more specific, are also created for each step in the process. The benefit of the parameterized approach to representing the problems is that these hints and feedback can also be written generically and then populated with specifics from the problem. For example, in a standard algebra problem type of ax+b=c, if a person subtracts the value of b from one side of the equation and not the other, the corrective feedback can be written as "You subtracted b from one side of the equation and not the other. You need to subtract b from both sides of the equation." This format allows for one general piece of feedback to be used in any problem of this type where the user makes this particular mistake.

Integrating AI Technology into Multimedia-based Instruction

Ultimately, the possible ways of integrating AI into multimedia-based instruction are perhaps limited only by the human imagination. We discuss two methods here, which were instantiated into technologies that were experimentally tested to determine their effectiveness.

The first method keeps the two technologies separate. In this method, students may learn a topic using multimedia technology such as a video or TV show and then practice what they learned in a separate AI-based system. This is perhaps the simplest form of integration and one that provides maximum flexibility since each component is then portable and reusable in other systems. It is also the easiest form to create from an architectural perspective since there is no need to worry about compatibility in programming languages, data structures, or operating systems.

This is the method used in the A-list Empire technology that combined AI with video-based instruction. In that example, the AI -based instruction was a standalone software product that could be paired with an existing video-based instruction system (in this case Khan Academy). The only "integration" issue is to make sure that the content for the two systems is aligned. This approach seems especially useful for a "learn then practice" teaching paradigm.

The second method actually embeds the AI directly into the multimedia platform. This is what was done in the interactive (AI) TV technology. Here, the student has the ability to interact directly with the TV characters during both the teaching and practice phases of the learning process. This method is more complex and less flexible, but may create a more powerful learning environment since it gives the student the opportunity to influence the delivery of the multimedia instruction through his or her responses.

Embedding AI into multimedia creates two main requirements. The first is technical: the AI must be created using a compatible architecture and programming language. For example, in the interactive TV technology, the core AI technology needed to be recoded into Action Script so that it could be embedded into Flash, which was the programming language used to create the animated TV show that the viewers saw.

The second requirement is to allow the AI to drive the multimedia instruction as needed. There are two components to this. First, a communication mechanism is needed between multimedia system and AI component. In this communication method, the multimedia system notifies the AI component of the user's actions, including questions the students ask or performance on practical problems. One method of doing this is through the use of "triggers" (Leddo, 1997) [12] that record what the student does and then translates that information into a form that is readable by the AI engine's knowledge base. The AI component then analyzes the actions to assess student mastery of the subject matter and learning needs. This analysis is used to select any modifications to the instruction, including providing corrective feedback.

This leads to the second component, a means to alter what the multimedia presents to the student based on the input of the AI component. If the alteration is a change to the flow of the content, as may be the case in a TV show or game, this can be accomplished either by having branches within the content flow as has been the case in classical interactive video training (cf., Schaffer and Hannafin, 1986) [16] or by real time rendering as often happens within games (cf., Stricker, 2013) [20]. In the present interactive (AI) TV technology, a branching format was used to simplify the number of paths a student could take within the TV show.

The other form of alteration to the flow of the TV show is to provide corrective instruction. In the present interactive (AI) TV technology used instructional sequences that were separately recorded with an eye to how they would be integrated into the point in the show where they would be called. When corrective instruction was needed, the flow of actions would be suspended and a TV character would deliver the needed instruction. Once the AI engine determined that the student learned what s/he needed to, the story line continued.

Accordingly, two versions of AI-based multimedia instructional technology were investigated. One was an embedded system used to provide interactive instruction within an Internet-delivered TV paradigm (developed by Education Online, Inc.) and the other was a standalone system (developed by A-list Empire) to piggyback on an Internet-delivered video-based instruction paradigm. The former used technology that was completely created from scratch. The latter was paired with instructional videos produced by Khan Academy.

Experiment 1

Methods

Participants: The experiment was conducted at Memorial Middle School in Spotswood, New Jersey, USA. There were 72 6^{th} grade students originally scheduled to participate in the study. Of these 36 were assigned to each group (interactive, non-interactive TV versions). Assignment was done by Spotswood School's superintendent based on students' 5th grade scores on New Jersey's ASK math assessment. Students were assigned so that each group was comprised of students with comparable math scores. Because of absenteeism, the final number of students was 33 in the interactive condition and 35 in the non-interactive condition.

Materials

The subject matter being taught was the mathematics topic of computing the number that changes a mean or average from one value to another. There were three types of materials used in the study: a pre-test, the TV shows themselves, and a post-test. The purpose of the pretest was to determine what pre-requisite knowledge participating students already had. The research team had been informed by Spotswood's superintendent that 6th graders had learned averages/means but had not learned how to compute an individual score that caused a previous mean to change its value to a new mean. In order to verify that the TV shows would be teaching new mathematics content, and therefore, could be considered responsible for any resulting performance, a two-problem pretest was constructed. The first problem gave the student seven numbers and asked them to find the mean or average of those numbers (both terms were used in case students were only familiar with one of those terms). The second problem gave students a table of a week's worth of cumulative average pitches thrown as of each day of the week and asked students to calculate how many pitches were thrown on Saturday. This problem had a format identical to the content taught in the TV shows.

The second kind of material used in the study was the two TV show versions: the non-interactive version and the interactive (AI) version. The third type of material was the post-test. The post-test contained questions of identical format taught in the TV shows. The first two post-test questions presented students with weekly tables of cumulative average soil temperatures and asked students to compute the soil temperature on Saturday. Question 3 had a similar format but involved number of trucks arriving at a warehouse, rather than soil temperature. Question 4 also had a similar format but involved number of classrooms visited by someone. At the end of the math questions, students were asked to rate the program they saw on a 4-point Likert scale with values very good, good, poor, very poor.

Procedure

The format of the study was as follows: a general project orientation, break up into groups, administration of a pretest, brief instruction on the user interface of the software (for the interactive (AI) condition), viewing the TV show, taking a post-test and then reconvening in a common room for a Q&A on the research.

The project orientation involved telling the students that they would be involved in a study evaluating new TV technology for the Internet. In the general orientation, students were not told of the different conditions. They were told that they would be asked to solve some math problems, they would watch a TV show on the computer that would teach them some math and then they would be given some problems to solve that used what they were taught. Students were encouraged to pay close attention to what they were taught on the TV show because they would take a test on the material afterwards even though they would not receive a school grade for the test.

After the orientation, students were split into three groups. The non-interactive group remained together and the interactive group was split into groups of 16 and 17 students to accommodate the school's computer facilities. Each group had an experimenter to give instructions and a middle school teacher to proctor the students' work.

Once the groups were split, each group was given the pretest. Students were asked to show all work. In order to avoid confusion, students were told that the tables they were given showed cumulative averages and not individual scores. They were given the example of grade point averages to illustrate what was meant by cumulative averages. They were reminded how a grade point average is the average of their tests to date and that each time they take a new test, the average changes. They were reminded that the problems were asking for the score that changed the average from one value to another just as a test score can change a grade point average from one score to another.

Once the pretest was completed, students used the software. In the interactive condition, students were given instructions on how to use the interactive parts of the software. The students were told that if they had questions on how to enter their work into the computer, they could ask the experimenter, but that the experimenter could not do the math for them. Students in the interactive TV condition were told that they could use calculators to help with the computations during the interactive part of the show and all students were told that they could use calculators during the post-test.

Students were then allowed to work at their own pace. In the non-interactive condition, two students watched the show on each computer. In the interactive condition, each student had his or her own computer. In all cases, students used headphones to listen to the show so that they would not be distracted by sound coming from other students' computers. Each room had an experimenter and a Spotswood teacher to proctor the students' work to insure there was no copying each other's work.

When students completed watching the show, they were given the post-test. Students were asked to show all work and were allowed to work at their own pace. Students turned in their work when they completed their tests to minimize the possibility of cheating. When all students completed their work, the three groups convened in the school's media room for a Q&A session.

Results

The primary purpose of the evaluation study was to test whether students using the interactive version of the TV technology would learn better than those using the non-interactive version. In order to insure that the comparison of these groups was valid, the pre-test scores were first analyzed. Question 1 of the pretest asked students to compute the average of seven numbers. Of the 35 students in the non-interactive group, 20 were able to solve this problem (57%). Of the 33 students in the interactive group, 13 were able to solve the problem (39%). This difference was not statistically significant. Question 2 of the pretest asked students to compute a daily soil temperature from a table of average soil temperatures (identical in format to the problem taught in the TV show). None of the students in either condition could solve that problem. Therefore, the pretest established that students learned something new in the TV show and that neither group had a significant advantage in baseline mathematics knowledge regarding averages.

The next step was to look at performance on the post-test. There were four questions in all. The first two mirrored the format and domains they had just learned (soil temperatures) and the second two had similar formats but different domains. All scoring of student responses was done by raters who were blind to which condition the students were in. In non-interactive TV condition, students, on average, correctly answered .29 questions out of four. In the interactive TV condition, students, on average, correctly answered 1.24 questions out of four. Given that the hypothesis was that the students in the interactive TV condition would score higher than those in the non-interactive TV condition, a one-tailed t-test was used to compare the two means. The difference was statistically significant, $t = 3.39$, $df = 66$, $p < .01$. However, the results would still be statistically significant if two-tailed tests were used.

Overall, students in the interactive TV condition performed about four times better than those in the non-interactive TV condition. These data suggest that interactive TV can produce larger learning gains than traditional TV where students merely watch the show.

Finally, students were asked to rate how much they liked the programs they saw. Choices were very poor, poor, good, very good. 97% of the students seeing the non-interactive version of the TV show and 88% of the students seeing the interactive version of the TV show rated the shows as "good" or "very good". This difference was not statistically significant. These ratings suggest that students can not only learn from interactive TV, but will enjoy the process as well.

Experiment 2

Methods

Participants: Participants were 20 students who were recruited from middle school and high schools in Fairfax and Loudoun counties in Virginia. Each one was enrolled in a geometry math class, which means that they had previously taken Algebra I, but had not taken Algebra II. It was necessary that each student had previously taken Algebra I since knowledge of the distributive property was necessary to learn the subject matter taught in the present study. However, it was also necessary that each student had not yet taken Algebra II since the subject matter of the present study, arithmetic operations with complex numbers, is a topic that is covered in the Algebra II curriculum. We wanted to make sure that participants in the present study had no prior knowledge of this topic. Each participated in the study without compensation.

Materials

The topic used in the present study was arithmetic operations (addition, subtraction, multiplication, and division) with complex numbers of the form a + bi, where i is the square root of -1. This topic is typically part of the Algebra II curriculum.

There were two core technologies used. First, for the control condition, there was the video-based lesson that teaches students how to perform arithmetic operations with complex numbers. This video-based lesson can be found on the Khan Academy website, www.khanacademy.org. It consists of seven videos, two that provide an introduction to complex numbers, one on complex conjugates and one each for addition, subtraction, multiplication, and division of complex numbers. The Khan Academy lesson software also includes a set of three practice problems per video (except for the introduction to complex numbers) and a space for entering the answer. Once the student answers an answer, there is a button, which, when clicked, states whether the answer is correct. If the student is unsure of how to do the problem, another button reveals hints, which students can use until they get the solution. These hints are not tied to the student's work but are general in nature. At no point does the student enter his or her work in solving the problem or is the student's work evaluated in any way by the Khan Academy software.

The AI software was taken from A-list Empire, www.alistempire.com, also presents videos that deliver general instruction on the topic followed by practice problems. The primary difference between the Khan Academy software and the A-list Empire software is that the latter includes an electronic worksheet that allows students to type in their work step-by-step. The worksheet is organized by lines, with one line given for each step. When a student is through typing in a step, s/he clicks on an enter button and the step is evaluated by the AI technology. If the step is correct, the student is notified in a feedback box below the worksheet. If the step is incorrect, the worksheet line the step is on is highlighted in yellow and the feedback box explains why the step is wrong and how the step is should be corrected. When the student completes the problem by entering the correct answer, the students is notified in the feedback box. As with the Khan Academy software, there is a hint button that students can use. In this case, the hints are tied to the step that the student has recently completed and gives the student information on how to complete the next step. There are three hints available, each at successive levels of detail. For example, in the problems involving division of complex numbers, the general hint tells the students to multiply by 1. The second hint tells the students to try to eliminate the imaginary part of the denominator. The third hint tells the student to multiply by the complex conjugate.

Procedure

The participants were first given a pre-test consisting of five problems to make sure that they did not already know the subject matter being taught in the experiment. The pre-test including problems in addition (one problem), subtraction (one problem), multiplication (two problems) and division of complex numbers (one problem). None of the participants had to be eliminated from the experiment because they already knew how to solve the pre-test problems.

Upon completion of the pre-test, participants were randomly assigned to experimental condition (Khan Academy software vs. A-list Empire software). As a result of the assignment, 10 participants wound up in the Khan Academy software condition and 10 participants wound up in the A-list Empire software condition. The first part of the instructional process was having participants in each group watch the Khan Academy videos on solving complex number problems using arithmetic operations. There were seven videos in all, two that were an introduction to complex numbers, one that taught addition of complex numbers, one that taught subtraction of complex numbers, one that taught multiplication of complex numbers, and two that taught division of complex numbers (one taught complex number conjugates and the other taught division of complex numbers). Since the main difference between the Khan Academy instructional software and the A-list Empire software is the use of AI to evaluate students' step-by-step work and provide hints and corrective instruction as needed, we wanted to keep the two conditions as close as possible. Using the same instructional videos in both conditions eliminates the possibility that any differences in resulting post-test performance could be attributed to differences in the instructional videos rather than the AI software.

After participants completed watching each instructional video (except the introduction), they were given three practice problems to solve for each type of arithmetic operation. In the Khan Academy software condition, participants were only able to enter answers to the practice problems. If they were stuck, they could press the button provided on the Khan Academy website to get general hints on how to solve the problem. These hints were not tied to the actual work done by the students, since the students had no way of entering their work on the Khan Academy website. In the A-list Empire software condition, students were given an electronic worksheet in which they entered their problem solving, step-by-step. If they were stuck, they would press the hint button and receive up to three hints as described in the educational technology section above. When they completed each step of the problem, they clicked on an enter button and would be notified if the step they entered was correct or would receive feedback on any mistake that they made. Participants then corrected the mistakes before moving on to the next step. The problem was considered complete once the student entered the correct answer.

When participants completed each instructional video and practice problem set, they were given a post-test. The post-test consisted of 20 problems of similar format to the ones taught in the Khan Academy videos. There were five questions each for complex number operations involving addition, subtraction, multiplication, and division. Participants were allowed no additional resources, such as calculators, to assist them in solving the problems. They were given scratch paper and pencils for computations. Also, to insure consistency, participants were not allowed to replay any of the videos.

Results

The answers to the 20 questions on the post-test were scored based on whether the correct answer was given. Because there were different types of problems based on the arithmetic operation involved, the data were broken out by both condition and problem type so that we could investigate whether there were any interaction effects as well as a main effect due to technology. The mean number of correct answers by participants, broken out by condition and problem type, is shown in Table 1. As can be seen in Table 1, participants in the Khan Academy software condition averaged 49.5% on the post-test. In US schools, this is generally considered to be a failing grade (F). Participants in the A-list Empire software condition averaged 90% on the post-test. In US schools, this is generally considered to be somewhere in the A grade range.

An analysis of variance was performed on the data and revealed a main effect of technology. The difference between the two means was statistically significant, $F(1,72) = 33.53$, $p < .0001$. This suggests that adding AI technology, as A-list Empire did, to a video-based e-learning system can greatly improve performance. There is a secondary finding that is worth noting in addition to the main effect. In any educational setting, there will always be some students who learn no matter how they are taught and some who will struggle. Therefore, in addition to looking at overall means, it is useful to explore how robust an educational technology is in teaching all of the students who use it. To do this, we examined the variability in scores between the two groups to see the degree to which the technology appears to help all students who use it.

In the Khan Academy group, the post-test scores ranged from 0 to 100, i.e., the full range of possible scores, suggesting that the technology is more effective for some students than others and that it is not particularly robust across students. Five of the Khan Academy students scored below 60 in the post-test (considered an F in most school districts) and with the exception of the one student who scored 100, no other student scored above 70 (considered in the C grade range in most school districts). In the A-list Empire group, the post-test scores ranged from 80 to 100. This suggested that, while some students did outperform others, in general, all students performed reasonably well. Of the 10 students in the A-list Empire condition, seven achieved scores of 90 or above (considered in the A grade range in most school districts) and the remaining three scored in the B grade range for most school districts. It is probably rare to find an educational intervention that produces performance in the A grade range for the vast majority of the students who use it and no performance lower than the B grade range.

In order to test statistically whether A-list Empire software is more robust across students than the Khan Academy software, we looked at the variability in performance. To do this, we conducted a Levene's Test for Homogeneity of Scores Variance across the eight different cells (2, technologies x 4 problem types). The results was statistically significant, $F(7,72) = 5.08$, $p < .001$, indicating that the A-list Empire software showed more consistent performance across students than did the Khan Academy software.

Another way to evaluate an educational technology's performance is to determine its robustness across the different types of content it teaches. Clearly, some content is easier for students to master than others, so it is no real achievement for an educational technology to claim that it can teach easy content. A review of the means in Table 1, along with a consideration of the procedures involved, suggests that some arithmetic operations involving complex numbers, such as division, are more difficult than others. This was confirmed by an analysis of variance, which revealed a main effect due to problem type $F(3,72) = 3.43$, $p < .05$.

Given that some problem types are more difficult than others, in order to test how robust the Khan Academy and A-list Empire softwares are with respect to problem type, we look at the technology by problem type interaction. An analysis of variance revealed that this interaction is significant, $F(3,72)=2.74$, $p < .05$. Looking at the individual problem types, we see a trend in the Khan Academy software condition for performance to drop off as students progressed from addition to subtraction to multiplication to division. In fact, the mean performance on division problems was statistically lower than it was for addition ($p < .01$) and subtraction ($p < .05$) problems. On the other hand, in the A-list Empire software condition, there was no significant difference in the mean performance across problem types. This suggests that the Khan Academy software works better on easier subject matter but poorly on complex subject matter (the mean performance in the division problems was 16%). In contrast, the A-list Empire software worked roughly equally well on both the easiest and most complex subject matter with mean performance ranging between 88% and 92%.

Table 1: Mean Number of Questions Answered Correctly Based on Question Type and Condition.

	Addition	Subtraction	Multiplication	Division	Total
Khan Academy	3.5	3.2	2.4	0.8	9.9
A-list Empire	4.5	4.6	4.5	4.4	18

General Discussion

The primary purpose of the research was to investigate whether adding an AI capability to multimedia-based instructional technology could enhance learning. Educational TV and video-based e-learning were chosen as platforms for investigation because these are two of the most prominent and popular forms of multimedia-based instructional technologies. Since there is no "brand name" leader in educational TV, we used software supplied by Education Online, Inc. for the versions of the interactive (AI) and non-interactive TV shows. Since Khan Academy is perhaps the most widely known and used video-based e-learning platform in the United States, it served as an ideal comparison to the AI-based counterpart, A-list Empire, which added AI technology to the video-based e-learning paradigm.

The results of both experiments show that the AI-based technology greatly outperforms their non-AI based counterparts. In the video-based instruction experiment, adding AI improved performance by 80% across all topics and by a factor of 5.5 in the hardest topic. The difference in performance amounted to the equivalent of mean student performance equating to an F grade in the typical public school system for students in the non-AI condition and to something in the A range for students in the AI condition.

Moreover, the AI technology showed itself to be robust across difficult educational topics and individual students, while the non-AI version showed a marked decline in student performance as subject matter got harder and had enormous variability in student performance. In the TV experiment, the difference in performance between AI and non-AI groups was greater than a factor of four. This is consistent with the results of the video-based instruction condition.

Earlier, we discussed two different ways of linking AI to multimedia instruction: pairing AI with multimedia as a separate module and embedded AI directly into the multimedia. It was suggested that embedded AI into multimedia might be more powerful. Unfortunately, the present experiments were not set up to do a direct comparison of standalone vs. embedded AI since both subject matter and student populations differed across experiments. However, we note that the standalone AI led to an 80% across the board increase in performance compared to the no-AI condition (although a ceiling effect certainly contributed to this), while embedded AI led to a fourfold increase in performance compared to the no-AI condition. Further research can help clarify whether embedded vs. standalone AI is a factor that matters and if so, under what conditions.

Conclusion

Members of the education community are always looking for ways to boost student achievement, particularly those that will be robust across both students and content. The AI technology reported in these experiments presents itself as a potential solution. The additional benefit of this technology is that it is relatively easy to implement and is scalable to the geographically disperse student population. Moreover, student ratings from Experiment 2 suggest that such AI-based technology is viewed highly favorably by the students who use it, suggesting that students would readily adopt the technology if used on a wide scale.

References

1. Anderson JR. The adaptive character of thought. Hillsdale, NJ: Erlbaum, 1990.
2. Aleven V, McLaren BM, Sewall J, Koedinger KR. The Cognitive Tutor Authoring Tools (CTAT): Preliminary Evaluation of Efficiency Gains. Human-Computer Interaction Institute. Pittsburgh, PA: Carnegie Mellon University, 2006.
3. Baloian NA, Pino JA, Hoppe HU. A teaching/learning approach to CSCL. Proceedings of the 33rd Hawaii International Conference on Systems Sciences, 2000.
4. Balslev T, De Grave WS, Muijtjens AMM, Scherpbier AJJA. Comparison of text and video cases in a postgraduate problem-based learning format. Medical Education, 2005; 39(11):1086-1092.
5. Boster FJ, Meyer GS, Roberto AJ, Inge C, Strom R. Some effects of video streaming on educational achievement. Communication Education, 2006; 55:46-62.
6. Carey J. *Television over the Internet.* Mahwah, NJ: Lawrence Erlbaum, 2003.
7. Claros I, Cobos R. An Approach for T-Learning Content Generation Based on a Social Media Environment. Proceedings of the 10th European Conference on Interactive TV and Video. New York, NY: Association for Computing Machinery, 2012, 157-160.
8. Donkor F. Assessment of learner acceptance and satisfaction with video-based instructional materials for teaching practical skills at a distance. The International Review of Research in Open and Distance Learning, 2011; 12(5):74-92.
9. Graesser AC, Conley M, Olney A. Intelligent tutoring systems. In K. R. Harris, S. Graham, & T. Urdan (Eds.), APA educational psychology handbook, Applications to learning and teaching. Washington, DC: American Psychological Association, 2012; 3:451-473.
10. Khan ML, Richards K, Wu ML. Understanding the Effectiveness of Video-based Instruction versus Text-based Instruction. In J. Herrington & C. Montgomerie Eds., Proceedings of EdMedia: World Conference on Educational Media and Technology. Association for the Advancement of Computing in Education (AACE). 2010, 3638-3646.
11. Kumar A, Kumar P, Basu SC. Student perceptions of virtual education: an exploratory study. Proceedings of the 2001 Information Resources Management Association International Conference, Toronto, Ontario, Canada, 2001.
12. Leddo J. Internet-based intelligent tutoring games. Proceedings of ED-MEDIA/ED-TELECOM 97. Charlottesville, VA: Association for the Advancement of Computing in Education, 1997.
13. Morgan PJ, Cleave-Hogg D, McIlroy J, Devitt JH. Simulation Technology: A Comparison of Experiential and Visual Learning for Undergraduate Medical Students. Anesthesiology, 2002; 96:10-16.
14. Piccoli G, Ahmad R, Ives B. Web-based virtual learning environments: a research framework and a preliminary assessment of effectiveness in basic IT skills training, MIS Quarterly, 2001; 25(4):401-426.
15. Roblyer MD, Edwards J. Integrating Educational Technology Into Teaching. Prentice Hall, Upper Saddle River, NJ, 2001.
16. Schaffer L, Hannafin M. The Effects of Progressive Interactivity on Learning from Interactive Video. Education and Communication Technology, 1986; 34:89-96.
17. Siegel PH, Omer K, Agrawal SP. Video simulation of an audit: An experiment in experiential learning theory. Accounting Education, 1997; 6:217-230.
18. Sottilare R, Graesser A, Hu X, Holden H. (Eds.) Design Recommendations for Intelligent Tutoring Systems: Learner Modeling. Orlando, FL: Army Research Laboratory, 2013, 1.
19. Steenbergen-Hu S, Cooper H. A meta-analysis of the effectiveness of intelligent tutoring systems on K–12 students’ mathematical learning. Journal of Educational Psychology, 2013; 105(4):970-987.
20. Stricker A. Slippery rock falls. [Software]. Loire: Virtual Harmony, 2013.
21. Van Tassel J. *Digital TV over Broadband.* New York, NY: Focal Press, 2001.
22. Zhang D, Zhou L, Briggs RO, Nunamaker Jr. JF. Instructional video in e-learning: Assessing the impact of interactive video on learning effectiveness. Information & Management, 2006; 43:15-27.