Optimizing Learning From Examples Using Animated Pedagogical Agents

Robert K. Atkinson
Mississippi State University

This study attempted to optimize a computer-based learning environment designed to teach learners how to solve word problems by incorporating an animated pedagogical agent. The agent was programmed to deliver instructional explanations either textually or aurally, while simultaneously using gaze and gesture to direct the learners to focus their attention on the relevant part of the example. In Experiment 1, learners presented with an agent delivering explanations aurally (voice plus agent) outperformed their control peers on measures of transfer. In Experiment 2, learners in the voice-plus-agent condition outperformed their peers presented with textual explanations on a variety of measures, including far transfer. In sum, an animated agent programmed to deliver instructions aurally can help optimize learning from examples.

A worked example is an instructional device that provides a model for solving a particular type of problem by presenting the solution in a step-by-step fashion. It is intended to provide the learner with an expert’s solution, which the learner can use as a model for his or her own problem solving. To date, in most experiments, worked examples have been visually fixed; that is, the examples simultaneously presented a problem and an expert’s solution steps. As such, these worked examples are similar to those found in traditional mathematics and science texts; however, instructional materials delivered on multimedia computer systems need not be limited in this way. For example, Stark (1999) and Renkl (1997) suggested that example processing can be enhanced by sequentially presenting problem states. According to Stark and Renkl, this type of presentation encourages learners to explain the examples to themselves by anticipating the next step in an example solution, then checking to determine whether the predicted step corresponded to the actual step—a phenomenon Renkl termed anticipative reasoning. According to Catrambone (1994, 1996, 1998), worked examples should be structured so they emphasize conceptually related solution steps (i.e., subgoals) by visually isolating them, by labeling them, or both. With regard to presenting examples that require learners to reference multiple sources of information, Mousavi and his colleagues (Mousavi, Low, & Sweller, 1995) offer a simple solution: Some segments of instructional information should be presented visually, whereas other segments should be presented aurally (i.e., mixed-mode format).

One advantage of using the computer to deliver instruction is that it enables instructional designers to combine multiple instructional principles or components in a worked example, which may prove to enhance its efficacy. According to Mayer’s (1997) generative theory of multimedia learning, computers—in contrast to a book-based medium—also provide a more favorable environment in which to implement some forms of effective instruction, such as the coordination of the visual presentation of sequential problem states with an auditory description of each of those states. For example, in Atkinson and Derry (2000), one way to structure an example within a computer-based multimedia environment so that learning can be maximized was to create a multicomponent worked example that (a) was sequential, in that it consisted of a sequential presentation of problem states; (b) was constructed to emphasize problem subgoals (i.e., it is subgoal oriented); and (c) incorporated a second modality that is coordinated with the sequential presentation of problem states (i.e., visually presented steps coupled with verbal instructional explanations). Learners exposed to these sequential, subgoal-oriented (SE–SO) examples with dual modes outperformed learners who were exposed to more traditional, simultaneous, non-subgoal-oriented examples on conceptually based measures of problem-solving transfer. Moreover, this difference occurred despite the fact that the examples in the latter condition were also dual mode.

Animated Pedagogical Agents

Although there are several computer-based tutoring systems currently under development, they fail to incorporate the features that are critical to successful human tutors, such as their motivational and affective features. Recently, one possible solution to this problem has begun to emerge in the form of animated pedagogical agents with lifelike qualities that operate within computer-based learning environments. In particular, it may be possible to structure an example-based learning environment so that a lifelike character can exploit verbal (e.g., instructional explanations) as well as nonverbal forms of communication (e.g., gaze, gesture) within the examples themselves in an effort to promote a learner’s motivation toward the task and his or her cognitive engagement in it. Lester and his colleagues (Johnson, Rickel, & Lester, 2000; Lester, Converse, Stone, Kahler, & Barlow, 1997; Lester, Stone, & Stelling, 1999) suggest that these lifelike characters are ideally suited to serve as tutors, coaches, or guides in knowledge-based learning environments. These agents reside in the learning environment by appearing as animated “humanlike” characters, which allows them to exploit nonverbal communication typically reserved for human–human interactions. For example, the agent can focus a learner’s
attention by moving around the screen, using gaze and gesture, providing nonverbal feedback, and conveying emotions.

Despite the potential of using animated pedagogical agents to emulate the actions of human tutors while operating in a computer-based learning environment, very little empirical research has been conducted on such agents to date. One of the principal contributors to this line of research is the IntelliMedia Initiative research program situated at North Carolina State University, which focused its efforts on creating problem-oriented learning environments inhabited by animated pedagogical agents. One of the initial studies produced by IntelliMedia researchers found that Herman the Bug, an animated pedagogical agent occupying a design-centered learning environment called Design-A-Plant, was capable of enhancing problem solving for middle school students learning botanical anatomy (Lester et al., 1997).

In a recently presented research report, Moreno and her colleagues (Moreno, Mayer, & Lester, 2000) used the Design-A-Plant learning environment described by Lester et al. (1997) to examine, in a well-controlled fashion, several issues related to the cognitive and affective impact of agents, including what the authors termed the personal-agent effect, image effect, modality effect, and dialogue effect. To examine for a personal-agent effect, that is, whether agents promote learning in a discovery-based learning environment, the authors compared a text-based version of the Design-A-Plant environment to one containing an animated agent (i.e., Herman the Bug). Moreno and her colleagues found that learners presented with an environment containing an animated pedagogical agent who communicated with the students using instructional conversations outperformed their peers in a comparable text-based environment on transfer problems, and they reported higher levels of motivation and interest. Searching for the source of this advantage across a series of experiments, the authors attributed this outcome to the agent’s voice (i.e., modality effect) and personalized language style (i.e., dialogue effect). Students learned better from “instructional conversations rather than non-personalized monologue-style communications” (Moreno et al., 2000, p. 744). The agent’s image (i.e., image effect), however, did not contribute to the personal-agent effect.

Overview of the Present Study

According to Atkinson and Derry (2000), one way to structure an example within a computer-based multimedia environment so that learning can be maximized is to create a worked example that is sequential, incorporates a second modality, and is subgoal oriented. Are there other ways to optimize the computer-enriched delivery method of examples? On the basis of the previous section, it appears worthwhile to examine whether a learning environment deploying an agent, one that coordinates aurally delivered instructional explanations with nonverbal cues to direct attention, is better than the same learning environment in which the explanations are text-based.

Another important issue that needs to be addressed is the effect of an agent’s image in the environment in which it is used to deliver explanations aurally. Specifically, does an agent’s image foster or hinder learning when it is programmed to explain additional visual information aurally—for instance, a set of problem states contained in an example—that appears on the screen simultaneously with the agent itself? Despite the empirical evidence that there is an increased cognitive capacity because the information is distributed across two working memories—verbal and visual working memories—as opposed to one, by using a dual-mode format in examples (Mousavi et al., 1995), the visual presence of the agent might be detrimental to learning by competing with the other visual information for the learner’s limited cognitive resources (Sweller, 1999). Indeed, Jeung and his colleagues (Jeung, Chandler, & Sweller, 1997) found that when the learning task involved visually complex, unfamiliar material, the superiority of dual-mode examples disappeared and only reappeared if the examples were restructured to include visual indicators—such as electronic flashing or animation—designed to draw attention to the visual segment of the example being alluded to by synchronized, aurally presented statements. Against this background, perhaps an agent could function as a visual indicator—comparable to the electronic flashing used by Jeung et al. (1997)—by using gesture (e.g., pointing) and gaze (e.g., turning head and eyes toward relevant material) to direct the learner’s attention to the pertinent material, thereby liberating cognitive resources that might otherwise have been devoted to searching the learning environment in order to connect the audio and visual information.

Thus, the purpose of the present study was to examine empirically how the efficacy of a computer-based learning environment could be improved through the use of an animated agent. While holding the instructional material and the learning environment constant, a pilot study and two experiments were conducted. The goal of the pilot study was to test the efficacy of an agent when coupled with a text-to-speech engine—a computer-based system able to read text aloud—to deliver instructional explanations aurally. Because the text-to-speech engine appeared to impede learning, it was replaced in Experiment 1 with recordings of a human voice. The goal of this experiment was to examine the impact of an agent’s visual presence, test the effect of the instructional explanation’s modality, and explore for a possible interaction between the visual presence of the agent and the modality of the explanations. The goal of Experiment 2 was to test for several effects, including an image effect, a modality effect, and an embodied agent effect. The latter effect is analogous to Moreno et al.’s (2000) personal-agent effect with one notable exception. In contrast to the personalized dialogue style used by Moreno et al.’s agent, the agent in the present study was designed to deliver monologue-style instructional explanations in a nonpersonalized fashion while using nonverbal cues such as gesture and gaze to direct attention. Both learning process and learning-outcome measures were collected. The learning-process measures included reported example difficulty and conceptual performance on practice problems. The learning-outcome measures included a posttest, which contained both near- and far-transfer items, and an affective questionnaire.

Pilot Study

Thirty undergraduate students were randomly assigned in equal proportions to one of the three conditions: voice plus agent, voice only, or text only (control). The text-only condition consisted of presenting learners with an example word problem followed by its sequentially presented solution steps, where each solution step was accompanied by a textual instructional elaboration designed to highlight what was occurring in that step (see Figure 1). In the
voice-only condition, a program designed to convert text into a computer-generated voice (i.e., text-to-speech engine) was used to deliver the instructional elaborations aurally. In the voice-plus-agent condition, an animated agent maintained a visual presence throughout instruction while delivering explanations aurally using the same computer-generated voice used in the voice-only condition (see Figure 2).

In this experiment, learners participated in two sessions. During the initial session, the participants filled out a demographic questionnaire, read through an eight-page review on solving proportion problems, and then completed an 11-item pretest. During the subsequent session, participants were asked to study the four example–practice problem pairs provided on the computer, with each pair consisting of a condition-specific worked example (i.e., voice-plus-agent, voice-only, or text-only control) along with a paired isomorphic practice problem. The learners were asked to provide the solution to each practice problem on a separate paper packet. Once they finished the four pairs of items, the participants were administered a posttest consisting of four multistep word problems—two near-transfer items and two far-transfer items.

Analyses of covariance (ANCOVAs) were used to analyze the three conditions (alpha of .05), on the learning-process measures (i.e., example difficulty and practice problem performance) and learning-outcome measures (i.e., near transfer, far transfer, and affective questionnaire). The pretest was used as the covariate and any significant differences uncovered by an ANCOVA were followed up with Fisher’s least significant difference (LSD) test, based on a familywise alpha of .05. Not only did this pilot study fail to establish any sort of effect of an animated agent, but the opposite result occurred on at least one learning-process measure: The scores of the participants in the text-only control condition were statistically superior to those of their counterparts in the voice-only condition in terms of practice problem performance, $F(2, 26) = 4.17, MSE = 6.90, p < .05$. Cohen’s $f$ statistic for these data yields an effect size estimate of 0.56, which corresponds to a large effect. These results were at odds with developing literature on animated pedagogical agents (Moreno et al., 2000). Of particular concern, however, was the absence of a general modality effect, inconsistent with dual-mode presentation techniques (e.g., auditory text and visual diagrams) producing typically superior effects compared with conventional, single-modality formats (e.g., visual test and visual diagrams) or visual-only formats (Moreno et al., 2000; Mousavi et al., 1995; Tindall-Ford, Chandler, & Sweller, 1997).
At least two factors could potentially have contributed to the lack of significance in this experiment: low power due to small sample size or a problem with the design of the instructional material. As a result of a series of informal exit interviews, it became apparent that the outcome of the pilot study could potentially be attributed to the voice used in the two conditions that relied on aurally presented instructional explanations, namely the voice-plus-agent and voice-only conditions. Unlike Moreno et al.’s (2000) animated agent, which relied on a human voice to deliver its instructional content, the agent in the pilot study used a voice generated by a text-to-speech engine that lacked many of the qualities associated with a human voice. A number of participants reported having to concentrate to understand what the computer-generated voice was saying. This outcome may have been the result of limitations associated with the text-to-speech engine’s voice quality, which can sound synthesized. To date, text-to-speech engines are unable to adequately capture the subtle nuances of a human voice, for instance the modulation and pace of a human voice. As a result, most text-to-speech voices sound synthesized, which makes them difficult to listen to for prolonged periods of time. Because the text-to-speech engine used in the pilot study could have compromised the performance of the voice-only and voice-plus-agent conditions that relied on aurally delivered explanations, the subsequent experiments were conducted using audio recordings of a human tutor in place of the text-to-speech engine.

Experiment 1

This experiment was designed to examine the effectiveness of an animated agent programmed to deliver monologue-style instructional explanations in a nonpersonalized fashion while using nonverbal cues such as gesture and gaze to direct attention. In particular, the focus was on whether an agent capable of delivering explanations either textually or aurally was more effective in supporting example-based learning than a conventional single-mode (e.g., text only) or dual-mode example (e.g., voice only). Thus, Experiment 1 addressed the following two questions: (a) Are worked examples, when coupled with the visual presence of animated pedagogical agents, more effective at promoting learning than examples without agents; and (b) are examples containing aural instructional explanations more effective at promoting learning than examples with textual explanations (i.e., modality effect)?
Method

Participants and design. Fifty undergraduate students (9 men and 41 women) from the Educational Psychology and Psychology Departments at Mississippi State University volunteered to participate in the study. The experiment consisted of a $2 \times 2$ factorial design with a control. The first factor was the modality of explanation (voice or text); the second factor was the presence or absence of an animated agent (agent or no agent). Thus, the participants were randomly assigned in equal proportions to one of the five conditions: voice plus agent, text plus agent, voice only, text only, or control, which did not contain any instructional explanations or an agent.

Pencil-paper materials. The pencil-paper materials included a demographic questionnaire, a review of proportion problems, a pretest, and an affective questionnaire. The materials used in this study consisted of instruction and problems involving multistep mathematical proportions. These materials were adopted with permission from Atkinson and Derry (2000), which Atkinson and Derry modified from a set originally developed by Derry, Weaver, Liou, Barker, and Salazar (1991). The questionnaire asked each learner to provide information (e.g., standardized test scores, number of post-secondary mathematics courses in progress or completed) that could be used to gauge the learner’s level of prior knowledge. The 10-page review of proportional reasoning covered such topics as ratios, proportional relationships, multiplicative relationships in proportional relationships, and identifying nonproportional relationships. It also described how to identify and solve proportional relationships in simple one-step word problems and included practice problems complete with answers for the learners to use in checking the accuracy of their problem solutions. The pretest consisted of 11 problems of varying difficulty. The pretest was designed to gauge the learner’s ability to solve, prior to treatment, proportion-word problems and to perform basic arithmetic operations in a variety of problem-solving contexts. It consisted of four one-step proportion problems, two multistep problems with one proportional relationship, three multistep problems with two proportional relationships, and two problems not involving proportional reasoning in their solutions. The following is an example of one of the multistep pretest problems with two proportional relationships:

Sheri, a student architect, wants to establish the difference in height between two buildings, the courthouse and the bank. If Sheri is 6-feet tall and casts a shadow 9-feet long and, at the same time, the shadows of the two building are 90- and 120-feet long, what is the difference in height between the two buildings?

In addition, this experiment incorporated a brief (five-item) pencil-paper affective questionnaire in which participants were asked to judge the effectiveness of the instructional program on a 5-point Likert-type scale. Specifically, the participants were asked to rate the instructional program that they had completed on interest (from boring to interesting) and comprehension (from very confusing to easily understood) and to indicate their agreement or disagreement with the following statements: “This is a good way to learn.” “I would like instruction in my courses like this.” and “I had trouble focusing my attention during this instruction.” The participants were instructed to respond with answers ranging from strongly agree to strongly disagree.

Computer-based learning environment. The computer-based learning environment used in this experiment was modeled after the Instructional Multimedia Modules (IMMs) designed to deliver worked-example instruction to learners learning to solve proportion-word problems in the experiment conducted by Atkinson and Derry (2000) and in Tutorials in Problem Solving (TiPS). TiPS consists of a problem-solving interface designed to help promote the learners’ ability to model and reason about story problems (Derry et al., 1994) and uses sequential, subgoal-oriented, computer-based worked examples as one of its instructional components. The IMM used in this study was created by Director 6.0 (Macromedia, 1997) software— an authoring tool for multimedia productions—and was designed for deployment within a Windows-based operating system. Regardless of which condition was being presented to a participant, the IMM contained a number of invariant structural features, including (a) its size, which was $800 \times 600$ pixels; (b) an instruction pane, for displaying the instructions for the current problem (see top left of Figure 1); (c) a problem text pane, for displaying the problem on which the worked example was based (see middle left of Figure 1); (d) a control panel, which allows the user to proceed through the instructional sequence at his or her own pace (see bottom left of Figure 1); and (e) a workspace for displaying the solution to the example’s problem (see right side of Figure 1).

The IMM included an instructional component, consisting of a total of four examples and four practice problems, and an assessment component, consisting of six posttest items. Specifically, the instructional component of the IMM consisted of four example–practice problem pairs, where each worked example was followed by an isomorphic practice problem. Across all of the conditions, the worked examples were structured—similar to the SE–SO examples used by Atkinson and Derry (2000)—to be sequential, that is, consisting of a sequential presentation of problem states, and to emphasize problem subgoals (i.e., subgoal oriented). That is, instead of appearing on the screen as a completely worked problem, as is the case with examples that simultaneously display all of the solution components (i.e., simultaneous examples), the sequential example appears initially unsolved. The learner then moves forward through the example and watches as problem states are successively added over a series of pages—similar to an animation, with the final page in the series representing the solution in its entirety. As illustrated by Figure 1, each example also contained two explicit cues designed to demarcate a problem’s subgoals: that is, each subgoal was visually isolated and labeled (e.g., “Total Amount 1”). The following is the cover story from one of the worked examples:

Bill’s Hometown Furniture Store creates custom-ordered furniture. Bill, the owner, received an order for 12 identical kitchen cabinets last week. Bill hired 4 carpenters to work for 5 days, and they made 7 cabinets in that time. However, 1 of the carpenters broke his arm over the weekend and, as a result, will be unable to help finish the order. If Bill has the 3 healthy carpenters complete the remaining cabinets, how long will it take them to finish the job?

The IMM was also configurable to run in one of five modes that reflected the five conditions of the present experiment. First, in the text-only condition (see Figure 1), learners were presented with an example word problem where each solution step was accompanied by a textual instructional elaboration designed to highlight what was occurring in that step (e.g., “Second, we need to set up another proportional relationship to determine the production time”). Second, the voice-only condition was identical in every respect to the text-only condition with one exception: the modality of the instructional explanations. Instead of using text-based instructional explanations in the examples, learners listened to a human tutor’s voice reading the textual explanations (the same voice as the one in the voice-plus-agent condition described below).

The voice-plus-agent (see Figure 2) and text-plus-agent (see Figure 3) conditions were indistinguishable from the text-only and voice-only conditions, respectively, with one notable exception: the presence of an agent. In these conditions, an animated agent maintained a visual presence throughout instruction, while explanations—the same explanations found in the text-only and voice-only conditions—were either delivered textually (text-plus-agent condition) or delivered aurally (voice-plus-agent condition). In the text-plus-agent condition, the explanations were delivered via a cartoon-like word balloon that appeared above the agent’s head. The word balloon was not present in the voice-plus-agent condition. The agent was also programmed to relate the textual or aural information (i.e., instructional explanations) with visual information (i.e., solution steps) by using nonverbal forms of communication. For instance, in Figure 3, the agent is gesturing and glancing toward an example’s solution step while...
using a word balloon to deliver the instruction explanation ("So, the travel package for John’s group will cost $18,947.09"). The agent conditions were created using Director 6.0 coupled with several off-the-shelf pieces of software: (a) Microsoft Agent, a collection of programmable pieces of software available from Microsoft, was used to support the presentation of the animated agent within this condition; and (b) XtrAgent 2.0, which enabled Director 6.0 (the program used to generate the original IMM) to animate the agent within the IMM. From the set of potential character files (i.e., agents) made available by Microsoft, the Peedy character file, an animated parrot, was selected for use in the present experiment. This particular agent was capable of 75 animated behaviors, including behaviors specifically designed to direct attention to objects on the screen, such as gesturing or looking in specific directions (e.g., up, down, left, right). Peedy was programmed to move around the work space, using gesture and gaze to highlight the example’s solution. Finally, in the control condition, the solution steps of each example were presented without instructional elaborations or the presence of an agent.

In every condition, after viewing the example problem, the learners were asked to rate the difficulty of the problem based on a 5-point Likert scale that ranged from 1 (very easy) to 5 (very difficult). After rating their understanding, a practice problem, parallel in structure to the example, was presented on the computer to the learners. The following is an example of a practice problem that is structurally isomorphic to the “Bill’s Hometown Furniture Store” example:

A local high school needs 120 classrooms painted over the summer. They hired 5 painters who worked for 6 days and completed 49 classrooms. Due to a conflict with management, however, 3 painters quit after the 6 days of work. If the 2 remaining painters finish the job, how long will it take them to finish painting the classrooms?

The learner was required to use the keyboard to enter a response to the practice problem before being given the final answer to the problem. The answers to the practice problems did not include solutions to problem steps nor any explanation about the solution.

The instructional component of the IMM was followed by the assessment component, in which a posttest was delivered that consisted of six multistep word problems, each one containing one or more proportional situations that were embedded in combination with other arithmetic operations. Of the six posttest items, three were near transfer and three were far transfer. The near-transfer items were structurally similar to items encountered during instruction. Although the surface story of each of the near-transfer items appeared dissimilar to the instructional items, they were each structurally isomorphic to one of the example–practice problem pairs. The following is an example of a near-transfer item that is structurally isomorphic to the “Bill’s Hometown Furniture Store” example:

Mike, a wheat farmer, has to plow 2,100 acres. He rented 6 tractors with people to drive for 3.75 days, and they completed 1,200 acres. He

Figure 3. Text-plus-agent condition.
must now finish the remaining acres. If he rents 4 tractor–drivers, how long will it take them to complete the plowing?

The remaining three items were far-transfer problems that were unlike any other problem seen during training. In particular, these multistep items varied from the example–practice problem pairs along a number of dimensions, including surface story and structural context (i.e., proportional situation embedded among different arithmetic operations), as well as number of proportion situations included. The following is an example of a far-transfer item:

Brian is selling newspapers at the rate of 3 newspapers every 10 minutes on one side of a downtown street, while Sheila—at her newsstand across the street—is selling papers at the rate of 8 newspapers every 20 minutes. If they decide to go into business together, how many newspapers will they sell in 40 minutes working at these rates?

Procedure. In this experiment, learners participated in two sessions that were not more than 1 week apart. During both sessions, the participants worked independently and were tested individually. For the majority of the participants, the first session lasted approximately 40 min. During the first session, the participants filled out a demographic questionnaire, then read through an eight-page review on solving proportion problems. The review also contained three basic problems that the participants were encouraged to try, followed by their complete solutions. When participants completed the review on proportion problems, they were given the 11-item pretest to complete.

During the second session, participants worked independently in a laboratory containing eight workstations. All participants wore headphones during the learning component of the IMM. Participants were asked to study the four example–practice problem pairs provided by the learning component of the IMM, with each pair consisting of a condition-specific worked example presented on the computer monitor along with a paired isomorphic practice problem presented on the monitor. The learners were asked to provide the solution to the practice problem on a separate paper packet. The learners had an opportunity to check the accuracy of their solutions using the IMM. Once they finished the four pairs of items, the participants were administered a six-item posttest by the assessment component of the IMM, which took approximately 30 min to complete. The test questions were presented individually on the computer screen, but the learners were asked to show their work on a separate paper packet. They were required to enter an answer before they could proceed to the next problem and they could not return to any of the preceding problems. The questionnaire was administered after the posttest.

Scoring. For each correct solution on the pretest, 1 point was awarded (no partial credit). The maximum score was 11 points. The protocols generated during problem-solving practice and on the posttest were coded for conceptual scores according to a set of guidelines for analyzing the written problem-solving protocols derived from research by Derry and her students (Atkinson & Derry, 2000; Derry et al., 1991; Tookey & Derry, 1994). These guidelines were designed to help gauge where the participant fell along a problem-comprehension continuum (see Table 1). According to these guidelines, each item was awarded a conceptual score, ranging from 0 to 3 points, depending upon the degree to which the participant’s solution was conceptually accurate. One research assistant who was unaware or “blind” to the condition independently coded each protocol. To validate the scoring system, two raters independently scored a random sample of 20% of the problem-solving protocols and agreed on scoring 92% of the time. Discussion and common consent was used to resolve any disagreement between coders.

The IMM also automatically coded one measure. The IMM recorded participants’ responses to one question posed after they exited each of the four examples. Participants were asked to rate the difficulty of the example that they had just seen on a scale of 1 (very easy) to 5 (very difficult). The participants’ responses to this query were summed across all four examples and divided by 4, thereby generating a measure of average perceived example difficulty, with values ranging from 1 to 5.

To create an overall affective score, the participants’ responses to the five questionnaire items were coded in the following manner. First, the participants’ rating of their interest in the program was scored on a scale of 1 (boring) to 5 (interesting) and their rating on comprehension was scored on a scale of 1 (very confusing) to 5 (easily understood). This was coupled with the participants’ responses to the other three questions on the questionnaire (e.g., “this is a good way to learn”) on a scale of 1 (strongly disagree) to 5 (strongly agree). The participants’ responses were summed across all five questions and divided by 5, thereby generating an average response on the affective measure, with values ranging from 1 to 5.

Table 1

<table>
<thead>
<tr>
<th>Points assigned</th>
<th>Scoring guideline</th>
</tr>
</thead>
<tbody>
<tr>
<td>3</td>
<td>Correct (ignoring minor computational–copying errors) There is perfect understanding of the problem and the student used a complete and correct strategy to arrive at an answer. The only error allowed is a simple calculation error (such as $3 \times 4 = 13$).</td>
</tr>
<tr>
<td>2</td>
<td>Substantial understanding of the problem The student basically understands the problem and is pursuing an identifiable solution strategy that is essentially correct. However, the student has made one or two fatal errors, such as mislabeling problem facts, confusing arithmetic operations, leaving out a small step, and so forth. On this type of error, a tutor probably would not have to intervene with full explanation or make a student start over but probably would help the student identify and correct error(s) in the current path.</td>
</tr>
<tr>
<td>1</td>
<td>Low understanding of the problem There is at least some slight evidence that one or more concepts underlying the problem are understood, although conceptual understanding of the problem is significantly flawed. In tutoring, a significant amount of explaining and coaching would be required in order to repair the problem solution, and the student would most likely have to start over in order to work the problem correctly (as opposed to making a repair in the current solution strategy).</td>
</tr>
<tr>
<td>0</td>
<td>No understanding exhibited There is no evidence of the student’s understanding the problem.</td>
</tr>
</tbody>
</table>
Results

Table 2 presents the means scores and standard deviations for each group on each of the dependent measures. The analysis consisted of two complimentary, alpha-controlled sets of analyses on both learning process and learning outcomes measures. First, comparisons between each of the “treatment” conditions and the control condition were made using Dunnett’s multiple comparison procedure based on a familywise alpha of .05. Then, a $2 \times 2$ (Agent $\times$ Voice) ANCOVA was conducted using the pretest as a covariate (alpha of .05). In the event a conditions main effect was found to be significant, pairwise comparisons among the various treatment conditions were examined using the sequential Hayter’s procedure, based on a familywise alpha of .05 (Seaman, Levin, & Serlin, 1991). Each measure was tested for homogeneity of regression and the results were found to be nonsignificant, $F < 1.2$. Because of the small sample sizes involved in this experiment, all of the following reported statistical conclusions were confirmed on the basis of small-sample exact permutation tests.

Analysis of learning-process measures. According to the first set of analyses that examined the relationship between each of the treatment conditions and the control condition (i.e., Dunnett’s multiple comparison procedure), voice-plus-agent participants reported significantly lower levels of perceived example difficulty than their control counterparts. Cohen’s $d$ statistic for these data yields an effect size estimate of 1.60 for reported example difficulty, which corresponds to large effect.

According to the second set of analyses, there was a significant main effect of voice on reported example difficulty, $F(1, 35) = 6.42$, $MSE = .67$, $p < .05$. Participants in the voice conditions reported that the examples were less difficult than the examples as reported by the participants in the no-voice conditions. Cohen’s $f$ statistic for these data yields an effect size estimate of 0.43 for reported example difficulty, which corresponds to a large effect. There was no significant main effect for voice on practice problem-solving performance, $F(1, 35) = .48$, $MSE = 8.07$, $p = .50$. There was also no significant main effect for agent on example difficulty, $F(1, 35) = 1.29$, $MSE = .67$, $p = .26$, or practice problem-solving performance, $F(1, 35) = .34$, $MSE = 8.07$, $p = .57$.

There was no significant interaction between the two factors (i.e., modality of explanation and the presence or absence of an animated agent) on perceived example difficulty, $F(1, 35) = 2.73$, $MSE = .67$, $p = .11$, or practice problem solving, $F(1, 35) = .01$, $MSE = 8.07$, $p = .92$.

Analysis of learning-outcome measures. According to the first set of analyses, the voice-plus-agent participants were statistically superior to the control participants on measures of both near and far transfer. The effect sizes were 1.04 on near transfer and 1.06 on far transfer, which both correspond to large effects.

The ANCOVA conducted on the measure of near transfer within the second set of analyses was significant for voice, $F(1, 35) = 5.91$, $MSE = 3.05$, $p < .05$. The participants who listened to the explanations outperformed their peers who had to read the explanations on near transfer. Cohen’s $f$ statistic for these data yields an effect size estimate of 0.42 for near transfer, which corresponds to a large effect. There was no significant main effect for voice on far transfer, $F(1, 35) = 3.12$, $MSE = 4.31$, $p = .09$. There was also no significant main effect for agent on near transfer, $F(1, 35) = .01$, $MSE = 3.05$, $p = .93$, or far transfer, $F(1, 35) = .13$, $MSE = 4.31$, $p = .72$.

There was no significant interaction between the two factors on near transfer, $F(1, 35) = .19$, $MSE = 3.05$, $p = .67$, or far transfer, $F(1, 35) = 1.01$, $MSE = 4.31$, $p = .32$.

Discussion

Are worked examples, when coupled with the visual presence of animated pedagogical agents, more effective at promoting learning than examples without agents? Participants in the voice-plus-agent condition reported lower levels of perceived difficulty with regard to the examples with which they were presented than did their counterparts in the control condition. Moreover, unlike their voice-only counterparts, the voice-plus-agent participants also outperformed their control peers on both near and far transfer. Thus, although the measurable effects were not as dramatic or as pervasive as might be expected, the findings provide a modicum of support for the conclusion that animated pedagogical agents are effective at promoting learning from examples.

Are examples containing aural instructional explanations more effective at promoting learning than examples with textual explanations (i.e., modality effect)? Participants who were exposed to the explanations aurally (i.e., voice conditions) reported less difficulty than did their peers who were exposed to the text-based

Table 2
Mean Scores and Standard Deviations by Condition for the Measures of Experiment 1

<table>
<thead>
<tr>
<th>Measure</th>
<th>Control</th>
<th>Text only</th>
<th>Voice only</th>
<th>Text + agent</th>
<th>Voice + agent</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$M$</td>
<td>$SD$</td>
<td>$M$</td>
<td>$SD$</td>
<td>$M$</td>
</tr>
<tr>
<td>Pretest</td>
<td>5.00</td>
<td>2.36</td>
<td>6.70</td>
<td>1.89</td>
<td>7.20</td>
</tr>
<tr>
<td>Reported example difficulty</td>
<td>3.45</td>
<td>0.79</td>
<td>2.85</td>
<td>0.80</td>
<td>2.55</td>
</tr>
<tr>
<td>Performance on practice problems</td>
<td>6.60</td>
<td>3.84</td>
<td>9.60</td>
<td>2.99</td>
<td>10.60</td>
</tr>
<tr>
<td>Posttest: far transfer</td>
<td>4.80</td>
<td>1.99</td>
<td>6.00</td>
<td>2.58</td>
<td>7.30</td>
</tr>
<tr>
<td>Affective scale</td>
<td>2.67</td>
<td>0.46</td>
<td>3.11</td>
<td>0.51</td>
<td>3.13</td>
</tr>
</tbody>
</table>

Note. Possible ranges of the variables: pretest (0–11), reported example difficulty (1–5), practice problems (0–12), near transfer (0–9), far transfer (0–9), and affective scale (1–5).
explanations on near transfer (large effect size). Moreover, the participants who were exposed to the explanations aurally (i.e., voice conditions) outperformed their peers who were exposed to the text-based explanations on the measure of near transfer on the posttest (large effect size). Thus, the findings are consistent with the supposition that aural instructional explanations are more effective at promoting learning than are examples with textual explanations. It is also noteworthy that simply replacing the text-to-speech engine used in the pilot study with human voice files permitted the general modality effect to reappear in this experiment. Perhaps the use of existing text-to-speech engines, in certain instances, may inhibit learning. An empirical examination of this issue certainly seems warranted.

Experiment 2

There were several limitations associated with Experiment 1 that may have contributed to the agent’s less-than-anticipated learning effects. One obvious limitation was the fact that the small sample sizes used in Experiment 1 limited the probability that the experiment would generate statistically significant results in favor of the animated agents (i.e., low power). Another shortcoming was the relatively small number of items incorporated into both the far- and near-transfer measures used in all of the experiments, which limited the reliability of these measures and, thus, may have inadvertently contributed to the problem of low power by introducing unnecessary error.

Experiment 2 was designed to overcome these limitations. One of the factors that contributed to the small sample sizes encountered in Experiment 1 was the reluctance on the part of potential volunteers to participate in an experiment that entailed two separate sessions. To address this problem, the pretest that was originally part of the initial session was eliminated in Experiment 2, and the entire experiment was collapsed into a single session. A more efficient, nonfactorial experimental design was also chosen for Experiment 2, one that permitted the removal of two conditions, thereby reducing the overall number of participants required to execute the experiment. To bolster the reliability of the posttest, the measure was expanded from six to eight items by adding one new near-transfer item and one new far-transfer item. The affective questionnaire was also replaced with one that included seven items.

With these modifications, Experiment 2 addressed the following questions: (a) Are worked examples, when coupled with the visual and auditory presence of animated pedagogical agents, more effective at promoting learning than more conventional text-only worked examples (i.e., embodied agent effect); (b) are worked examples, when coupled with the visual and auditory presence of animated pedagogical agents, more effective at promoting learning than more conventional voice-only worked examples (i.e., image effect); and (c) are two sensory modes better than one (i.e., modality effect)?

**Method**

**Participants and design.** Seventy-five undergraduate students (22 men and 53 women) from the Educational Psychology and Psychology Departments at Mississippi State University volunteered to participate in the study. The participants were randomly assigned in equal proportions to one of the three conditions: voice plus agent, voice only, or text only.

**Pencil-paper materials.** All of the materials were the same as those used in Experiment 1, with two exceptions: the posttest and the affective questionnaire. In an effort to enhance the reliability of the measure, the original posttest was expanded to include an additional two problems, one near transfer and one far transfer, for a total of eight items. The new near-transfer problem was similar in structure to an example–practice problem pair whereas the other new far-transfer problem required knowledge of proportions but was not explicitly similar in format to any of the instructional problem pairs.

The affective questionnaire from Experiment 1 was replaced with a similar measure that consisted of seven statements to which the participants were expected to respond on a 5-option Likert-type scale from I disagree to I agree. The statements included (a) “I have learned to solve proportion problems based on this instruction,” (b) “Learning was fun,” (c) “I would prefer learning from such examples and problems when I have to study ‘mathematized’ contents next time,” (d) “I felt curious,” (e) “The examples and problems helped me to understand the application of the proportions,” (f) “I was interested in learning about proportions,” and (g) “The examples and problems were well designed.”

**Computer-based learning environment.** The IMM used in Experiment 1 was modified for the purposes of the second experiment. Specifically, the IMM was modified to deliver only three of the five conditions available in the previous experiment (i.e., text only, voice only, and voice plus agent). In the text-only condition, as in its counterpart in Experiment 1, learners read the textual explanations without an agent being present (see Figure 1). In the voice-only condition, learners listened to the human tutor’s voice (the same voice used in the voice-plus-agent condition) highlighting solution steps, but no agent was present. This condition was identical to Experiment 1’s voice-only condition. In the voice-plus-agent condition, learners saw an agent highlighting solution steps both verbally—with a human tutor’s voice (the same voice used in the voice-only condition)—and nonverbally—by using gesture and gaze (see Figure 3). This condition was identical to Experiment 1’s voice-plus-agent condition.

**Procedure.** Unlike in Experiment 1, learners participated in a single session in this experiment. During this session, participants worked independently in a laboratory containing eight workstations. During this session, the participants filled out a demographic questionnaire, then read through an eight-page review on solving proportion problems. The review also contained three basic problems that the participants were encouraged to try, followed by their complete solutions. When participants completed the review on proportion problems, they were asked to study the four example–practice problem pairs provided by the learning component of the IMM, with each pair consisting of a condition-specific worked example presented on the computer monitor along with a paired, isomorphic practice problem presented on the monitor. As in Experiment 1, the learners were asked to solve the practice problem on paper. The learners were permitted to check the accuracy of their solutions using the IMM. After the instructional component concluded, the participants were administered an eight-item posttest by the assessment component of the IMM, which took approximately 50 min to complete. As in Experiment 1, the test questions were presented individually on the computer screen, but the learners were asked to show their work on paper. Before they could proceed to the next problem, the learners were required to enter an answer. The affective questionnaire was administered after the posttest.

**Scoring.** Scoring was identical to that in Experiment 1 except on the affective questionnaire where the participants’ responses were summed across all seven questions and divided by 7, thereby generating an average response on the affective measure with values ranging from 1 to 5.

**Results**

Table 3 presents the mean scores and standard deviations for each group on each of the dependent measures. An analysis of
variance was conducted on each learning-process measure and performance measure (alpha of .05). Significant main effects were followed up with Fisher’s LSD test, based on a familywise alpha of .05.

Analysis of learning-process measures. For reported example difficulty, means of the three treatments differed statistically from one another, $F(1, 72) = 4.91, MSE = .63, p < .05$. The participants in the voice-plus-agent condition outperformed the participants in the text-only condition (Fisher’s LSD test). There was also an overall difference among the conditions for performance on practice problems, $F(1, 72) = 3.11, MSE = 14.03, p = .05$. The participants in the voice-plus-agent condition were statistically superior to their peers in the text-only condition in terms of the average performance on practice problems. Cohen’s $d$ statistic for these data yields effect size estimates of 0.37 for reported example difficulty and 0.29 for average performance on practice problems, which corresponds to a medium-to-large effect and medium effect, respectively.

Analysis of learning-outcome measures. There was a significant conditions main effect for the near-transfer items on the posttest, $F(1, 72) = 3.21, MSE = 12.01, p < .05$. The near-transfer performance of the participants in the voice-plus-agent condition was statistically superior to that of their peers in the text-only condition (Fisher’s LSD test). Moreover, there was a significant conditions main effect for far transfer, $F(1, 72) = 3.57, MSE = 13.30, p < .05$. Participants assigned to the voice-plus-agent condition were statistically superior to their counterparts in both the voice-only and text-only conditions in terms of far-transfer performance. The three treatments also differed statistically from one another on the affective questionnaire, $F(1, 72) = 4.63, MSE = .46, p < .05$. The participants in the voice-plus-agent condition outperformed the participants in both the voice-only and text-only conditions. Cohen’s $f$ statistic for these data yields effect size estimates of 0.30 for near transfer (medium effect), 0.31 for far transfer (medium effect), and 0.36 for the affective questionnaire (medium-to-large effect).

Supplemental analysis. Unlike in Experiment 1, the participants in the voice-only condition in the present experiment did not outperform their peers in the text-only condition on any of the learning process or learning-outcome measures (i.e., no modality effect). In an effort to explore this further, the two voice conditions were combined and compared with the text-only condition—contrast: ([voice plus agent] + [voice only]) vs. text only—across the measures, each at alpha of .05. With findings similar to those of Experiment 1, the participants in the voice conditions reported less perceived difficulty ($M = 2.37$), $t(72) = 2.76, SE = .19, p < .05$, and higher near-transfer scores ($M = 7.78$), $t(72) = 2.24, SE = .85, p < .05$, than did the participants in the text-only condition ($M = 2.87$ and $M = 5.88$, respectively). Cohen’s $d$ statistic for these data yields effect size estimates of 0.65 for perceived example difficulty and 0.53 for near transfer, which both correspond to medium effects. Unlike in Experiment 1, the participants in the voice conditions ($M = 7.56$) also outperformed their text-only counterparts ($M = 5.64$) on practice problem-solving performance, $t(72) = 2.09, SE = .92, p < .05$. The voice condition participants ($M = 3.59$) also reported a higher average affective score than their text-only peers ($M = 3.26$), $t(72) = 2.03, SE = .17, p < .05$. Cohen’s $d$ statistic for these data yields effect size estimates of 0.49 for perceived example difficulty and 0.48 for average performance on practice problems, which both correspond to medium effects.

Discussion

Are worked examples, when coupled with the visual and auditory presence of animated pedagogical agents, more effective at promoting learning than more conventional text-only worked examples (i.e., embodied agent effect)? Adding an animated agent to a learning environment for teaching proportion-word problems fostered learning on a variety of measures. Specifically, during instruction, participants presented with an agent produced more conceptually accurate solutions on practice problems and perceived the worked-out examples to be less difficult than did their counterparts who were presented with textual explanations alone. After instruction, the participants generated more conceptually accurate solutions on near- and far-transfer items than did their text-only peers. The agent participants also responded more favorably to the learning environment on an affective measure as indicated by their higher average response to the affective questionnaire. Overall, these results indicate that an agent delivering monologue-style instructional explanations while using nonverbal cues (e.g., gesture, gaze) to direct a learner’s attention is more effective at fostering learning—as indicated by improved learning when the learning environment is supplemented with an agent—

Table 3

<table>
<thead>
<tr>
<th>Measure</th>
<th>Text only</th>
<th>Voice only</th>
<th>Voice + agent</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$M$</td>
<td>$SD$</td>
<td>$M$</td>
</tr>
<tr>
<td>Reported example difficulty</td>
<td>2.87</td>
<td>0.85</td>
<td>2.50</td>
</tr>
<tr>
<td>Performance on practice problems</td>
<td>5.64</td>
<td>3.60</td>
<td>6.84</td>
</tr>
<tr>
<td>Posttest: near transfer</td>
<td>5.88</td>
<td>4.00</td>
<td>7.20</td>
</tr>
<tr>
<td>Posttest: far transfer</td>
<td>3.92</td>
<td>3.41</td>
<td>4.32</td>
</tr>
<tr>
<td>Affective scale</td>
<td>3.26</td>
<td>0.76</td>
<td>3.38</td>
</tr>
</tbody>
</table>

Note. Possible ranges of the variables: reported example difficulty (1–5), practice problems (0–12), near transfer (0–12), far transfer (0–12), and affective scale (1–5).
than is a comparable text-based learning environment, which is consistent with an embodied agent effect.

Are worked examples, when coupled with the visual and auditory presence of animated pedagogical agents, more effective at promoting learning than more conventional voice-only worked examples (i.e., image effect)? The results of Experiment 2 are consistent with an image effect, at least in terms of fostering far transfer. As suggested, the agent appeared to function as a visual indicator—akin to the electronic flashing used by Jeung et al. (1997)—by using gesture and gaze to guide learners’ attention to the relevant material. These nonverbal cues (e.g., gesture, gaze) apparently did not overburden the learners’ limited cognitive resources (Sweller, 1999)—as indicated by improved learning when the agent’s image was present. Perhaps the agent’s use of nonverbal cues enabled the learners to dedicate their limited cognitive resources to the task of understanding the underlying conceptual segments of the worked-out examples. Without the benefit of the agent’s image, perhaps the voice-only participants were occupied with searching the learning environment in order to connect the audio and visual information, which prevented them from committing their restricted cognitive resources to the task of understanding the deep structure of the example at hand.

Are two sensory modes better than one (i.e., modality effect)? Although in the initial analysis, the participants in the voice-only condition did not outperform their peers in the text-only condition, the results of the supplemental analysis support the modality effect. Specifically, in contrast to their counterparts who were exposed to textual explanations, participants who listened to a human tutor’s voice deliver the instructional explanations perceived the examples presented during instruction to be less demanding. They also produced more conceptually accurate solutions on the practice problems and the near-transfer items on the posttest.

General Discussion

In sum, it is apparent from the evidence compiled in the present study that learners who are engaged in mathematical thinking can benefit on a variety of cognitive and affective measures by working within a learning environment that contains an animated pedagogical agent—in particular, an agent capable of delivering instruction aurally and using forms of nonverbal communication to support learning. The type of animated agent used in the present study appears capable of optimizing a computer-based multimedia environment for learning from worked-out examples by fostering learning, as indicated by improved performance on both learning-process measures and learning-outcome measures.

The present findings provide additional support for previously published research. First, this study supports the previously mentioned research of Moreno and her colleagues (Moreno et al., 2000) on the efficacy of animated agents operating in multimedia learning environments, despite the fact that the studies differed in two distinct ways: (a) the agent used in the present study used what Moreno et al. would characterize as a “non-personalized monologue-style” of communication with the learners, a style of communication that, according to their research, did not enhance learning relative to the more personalized dialogue style their agent used (i.e., dialogue effect); and (b) the present study used an example-based learning environment for learning proportional reasoning whereas Moreno and her colleagues used a discovery-based learning environment for learning botanical anatomy. This study also provides additional empirical support for the proposal that example processing and problem-solving performance can be improved by incorporating the use of a dual mode of presentation in example-based instruction (Mousavi et al., 1995) and, more generally, that the use of a dual mode of presentation enhances learning outcomes in multimedia learning environments (Mayer, 1997; Moreno et al., 2000).

Finally, the results of this study suggest a number of avenues for future research, some designed to address shortcomings of the present study, as well as others that direct attention to several productive research questions raised in the present study. Foremost among these recommendations is the need to replicate the findings discussed herein. In particular, the findings from Experiment 2—including the embodied agent effect and image effect—need to be empirically replicated in order to strengthen the claim that animated pedagogical agents support learning in example-based learning environments. Future research should also attempt to empirically establish differences between the effectiveness of text-to-speech engines versus audio recordings of human tutors. Finally, although these findings might generalize to other areas of mathematics, research should examine whether these findings generalize across other semantically rich domains.

References


Received September 25, 2000
Revision received August 14, 2001
Accepted August 14, 2001