Journal of Educational Psychology
2005, Vol. 97, No. 1, 117–128

0022-0663/05/$12.00 DOI: 10.1037/0022-0663.97.1.117

Role of Guidance, Reflection, and Interactivity in an Agent-Based Multimedia Game

Roxana Moreno
University of New Mexico

Richard E. Mayer
University of California, Santa Barbara

The authors investigated whether guidance and reflection would facilitate science learning in an interactive multimedia game. College students learned how to design plants to survive in different weather conditions. In Experiment 1, they learned with an agent that either guided them with corrective and explanatory feedback or corrective feedback alone. Some students were asked to reflect by giving explanations about their problem-solving answers. Guidance in the form of explanatory feedback produced higher transfer scores, fewer incorrect answers, and greater reduction of misconceptions during problem solving. Reflection in the form of having students give explanations for their answers did not affect learning. Experiments 2 and 3 showed that reflection promotes retention and far transfer in noninteractive environments but not in interactive ones unless students are asked to reflect on correct program solutions rather than on their own solutions. Results support the appropriate use of guidance and reflection for interactive multimedia games.

Do educational games have the potential for improving academic learning? For example, Prensky (2001, pp. 4–5) proposed that "by marrying the engagement of games and entertainment with the content of learning and training, it is possible to fundamentally improve the nature of education and training for . . . students and trainees." The enthusiasm for educational games rests in their potential to motivate and engage learners, but there has not been adequate research on how to design game environments so that they foster deep understanding in learners (de Jong & van Joolingen, 1998; Jonassen, 1996). The goal of the present set of studies is to pinpoint the role of guidance and reflection in promoting scientific understanding in agent-based multimedia games. Concerning guidance, we are interested in the role of explanatory feedback, in which a pedagogical agent provides principle-based explanations for correct answers during a problem-solving session. Concerning reflection, we are interested in the role of a simplified version of elaborative interrogation, in which a pedagogical agent asks the learner to provide an explanation for an answer during a problem-solving session. Moreover, we are interested in examining whether guidance and reflection have the same affective and learning effects for two different multimedia environments: interactive, in which the pedagogical agent asks the learner to participate in the game by selecting an answer for a problem, and noninteractive, in which the learner receives from the pedagogical agent an answer for a problem.

This research was supported by National Science Foundation Grant 0238385 entitled "Guided Interactive Virtual Environments for Case-Based Learning." Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the National Science Foundation.

Correspondence concerning this article should be addressed to Roxana Moreno, Educational Psychology Program, University of New Mexico, Albuquerque, NM 87131-1246, or to Richard E. Mayer, Department of Psychology, University of California, Santa Barbara, CA 93106. E-mail: moreno@unm.edu or mayer@psych.ucsb.edu

For example, consider an environmental science simulation game presented on a desktop computer in which the learner goes on a space ship to a new planet. On the learner's arrival, an on-screen agent named Herman-the-Bug explains that the planet has certain environmental conditions—such as heavy rainfall and strong winds—and that the learner's job is to design a plant that will survive. Each learning trial consists of up to seven parts: (a) the agent asks the learner to click on an answer (interactivity step); (b) the agent asks the learner to generate an explanation for the answer (directed reflection step); (c) the agent tells the learner whether the answer is correct (feedback step); (d) the agent provides an explanation for the answer (guidance step); (e) if students click on the right answer, the agent moves on to Step g; (f) if students did not click on the correct choice for a plant part, the agent asks them to try to design the plant part again; and lastly (g) the agent shows students the right choice and moves on to the next step.

In the interactivity step, the learner is asked to select the type of roots, stem, and leaves that are best suited for the plant's survival on the planet. For example, a set of eight roots is presented, and the learner clicks on one of them. We manipulated the level of interactivity by allowing students to select the appropriate answer (interactive treatment) or by having the pedagogical agent select the correct answer (noninteractive treatment). In the directed reflection step, the learner is asked to give an explanation for the type of root, stem, and leaves that he or she designed. We manipulated the level of directed reflection by either asking or not asking students to give an explanation for their answer. Finally, in the guidance step, the pedagogical agent provides a principle-based explanation of why a certain plant design is appropriate for a certain environment. We manipulated the level of guidance by either providing or not providing the agent's explanation of an answer.

Table 1 summarizes how we implemented each of the four general instructional concepts used in the game: *interactivity*, which we implemented as asking the learner to give a solution to

Table 1
How Four Instructional Concepts Were Implemented and the Cognitive Processes They Were Intended to Prime

General concept	Specific implementation	Step	Cognitive process
Interactivity	Problem solving: Learner is asked to give an answer to a problem	a	Organizing and integrating
Reflection	Elaborative interrogation: Learner is asked to explain why an answer is correct	b	Organizing and integrating
Feedback	Corrective feedback: Learner is told whether an answer is correct	c	Selecting (partial)
Guidance	Explanatory feedback: Program explains why an answer is correct	d	Selecting

a problem (such as Herman-the-Bug describing a specific environment and asking the learner to choose the appropriate roots, stem, and leaves for a plant to survive); *reflection*, which we implemented as asking the learner to explain why the answer is correct (such as Herman-the-Bug asking the learner to justify the answer); *feedback*, which we implemented as Herman-the-Bug telling the learner whether the selected answer is correct; and *guidance*, which we implemented as having Herman-the-Bug explain why the selected answer is or is not correct.

Our research is guided by the following question: To what extent do reflection, guidance, and interactivity within a multimedia program promote the learner's meaning making? To answer this question, we first present and exemplify a general model for describing the cognitive processes involved in meaning making. Second, we review some instructional methods that are intended to foster deep learning. Third, in three experiments, we tested an application of such methods within the scenario of an agent-based multimedia environment. Finally, in light of the results, we review the roles of reflection, guidance, and interactivity in order to derive instructional design principles for agent-based multimedia learning.

A Cognitive Model of Multimedia Learning

We begin with the idea that meaningful learning occurs when a learner actively constructs a coherent knowledge representation in working memory. For example, in our agent-based computer game about environmental science, a learner may attempt to build a mental model of the relationship between the design of plants and the weather conditions by processing the respective visual and verbal representations. According to a cognitive model of multimedia learning, the role of an instructional designer is to create environments in which the learner interacts meaningfully with multimedia materials, including fostering the cognitive processes of selecting relevant information, organizing that information into coherent representations, and integrating these representations with existing knowledge (Mayer, 2001; Mayer & Moreno, 2002; Moreno & Mayer, 2002b).

How can multimedia games foster the cognitive processes of selection, organization, and integration in learners? Unlike theories of learning that emphasize the addition of presented information to long-term memory, our cognitive model of multimedia learning focuses on the way that knowledge is constructed by the learner in working memory. On the basis of the main assumptions of the model, we propose that meaningful learning depends on two conditions being met: First, the learner must engage in the cognitive process of selecting relevant aspects of the images and words included in the multimedia presentation. For example, learners need to pay attention to the different weather conditions and plant features that may apply to those particular weather conditions. The right column of Table 1 shows that guidance from a pedagogical agent is intended to help meet this condition by directing the learner's attention to relevant information. Second, the learner must activate relevant prior knowledge, integrate the incoming material with that prior knowledge, and organize the incoming material in a coherent structure. For example, learners need to try to make sense of the correct plant design for each weather condition so that they can eventually construct a mental model that will allow them to infer the plant designs for any given environment. We refer to these cognitive processes as *integrating* and *organizing*. The right column of Table 1 shows that interactivity and reflection are intended to help meet this condition.

Meaningful learning—in which learners both select relevant information and also organize and integrate it with existing knowledge—is more likely to occur when (a) learners receive guidance (to help them select the appropriate material) and (b) engage in reflection and/or interactivity (to help them mentally organize and integrate the material). Meaningful learning is less likely to occur when learners engage in reflection and/or interactivity but without guidance (i.e., organizing and integrating without selecting).

Instructional Methods to Foster Multimedia Learning

In this section, we emphasize three aspects of instruction that can affect how knowledge is learned: guidance, reflection, and interactivity. First, in terms of guidance, one of the choices that an instructional designer needs to make concerns the amount of support the learning environment will provide to ensure that students are building knowledge successfully. In a discovery-based multimedia game where students are allowed to discover knowledge about the task at hand, the cognitive role of a pedagogical agent may consist of giving more or less feedback on students' interactions with the program. For example, in guided discovery environments, the agent may guide or scaffold the process of knowledge construction by providing explanatory feedback on students' choices to facilitate the cognitive process of selection (Moreno, 2004). On the other hand, in pure discovery environments, the role of the agent may be limited to providing students with the minimum amount of information (i.e., telling whether the student's response is correct). Although, it may be argued that discovery methods facilitate active learning by allowing students to explore, manipulate, and test hypotheses (Bruner, 1961; Gagne, 1965; Wittrock, 1966), when students learn science in classrooms with pure-discovery methods and minimal feedback, they often become lost and frustrated, and their confusion can lead to misconceptions (Brown & Campione, 1994; Hardiman, Pollatsek, & Weil, 1986; Mayer, 2004). This is particularly important in the case of novice learners, who lack proper schemas to guide them in the selection of relevant new information (Tuovinen & Sweller, 1999). In the present studies, we focus on a specific type of guidance that can be called *explanatory feedback* in which the animated pedagogical agent explains why a certain plant design is or is not correct. As shown in Table 1, we expected this form of guidance to facilitate the cognitive process of selecting.

A second instructional method embedded in our educational game is reflection, which we implemented as a form of elaborative interrogation. This method has consistently been shown to improve students' learning from text, as measured by retention and comprehension tests (Bruning, Schraw, & Ronning, 1999). Pressley and colleagues (Martin & Pressley, 1991; Pressley, McDaniel, Turnure, Wood, & Ahmad, 1987; Pressley, Symons, McDaniel, Snyder, & Turnure, 1988; Willoughby, Waller, Wood, & MacKinnon, 1993; Woloshyn, Paivio, & Pressley, 1994; Wood, Pressley, & Winne, 1990) improved learning by asking students to answer "why" questions about information they have just read. Seifert (1993) reviewed research showing that when students use elaborative interrogation to process expository paragraphs, their comprehension is enhanced by stimulating the inferences made about the textual content. Chi, de Leeuw, Chiu, and La Vancher (1994) have summarized similar positive effects on learning from text when students are requested to provide self-explanations during reading. Thus, we focused on a specific type of reflection in which learners are asked to explain correct answers. As shown in Table 1, we expected this form of reflection to prime the cognitive processes of organizing and integrating.

Third, concerning interactivity, we are interested in whether students learn more deeply when they are asked to produce answers rather than to simply receive the correct answer from the pedagogical agent. Interactivity may activate some of the cognitive processes required for meaningful learning—such as activating prior knowledge in long-term memory and trying to organize and integrate it with incoming information (Anderson & Pearson, 1984; Doctorow, Wittrock, & Marks, 1978; Moreno, Mayer, Spires, & Lester, 2001). However, if the learner fails to find the correct answer and an explanation for the correct answer (i.e., selecting), meaningful learning cannot occur. Although interactivity is a central feature of discovery learning environments, prior research has shown that discovery environments may be improved with some instructor-based guidance (de Jong & van Joolingen, 1998; Mayer, 2004; Moreno, 2004). We focused on a specific type of interactivity, in which learners are asked to choose solutions to problems. As shown in Table 1, we expected this form of interactivity to prime the cognitive processes of organizing and integrating.

In sum, in this research, we focused on three instructional methods intended to promote meaningful learning in an agent-based multimedia game: providing students with guided explanations about their choices (i.e., guidance), promoting students' reflection by asking learners to justify their answers (i.e., reflection), and encouraging students to actively construct answers to problems posed by a pedagogical agent (i.e., interactivity). We focused on two measures of learning: retention, in which we assessed memory for the basic factual information that was presented, and problem-solving transfer, in which we asked students to solve new problems based on the principles learned in the multimedia program. In addition, to examine students' impressions about the multimedia program, we included a program-rating questionnaire in which we asked students to rate how interesting, helpful, friendly, understandable, and easy to learn the program was. We expected the highest performance on retention and transfer measures when the instructional methods prime both the process of selecting (via guidance) and the processes of organizing and integrating (via reflection or interactivity). Finally, although a cognitive theory of multimedia learning does not offer specific predictions on students' affective measures, a goal of this set of studies was to examine whether the instructional methods embedded in our program produce positive effects on students' program ratings. The program rating was intended to provide preliminary information concerning the learners' interest in the lessons, as a supplement to our major focus on cognitive learning outcomes.

Experiment 1

The purpose of Experiment 1 was to contribute to a cognitive theory of multimedia learning by examining how guidance and reflection affect learning in an interactive multimedia game. In Experiment 1, all learners received the interactivity treatment (Step a) and feedback concerning the correct answer (Step c). We varied reflection, that is, some learners were asked to explain the answer they selected (Step b) and some were not (No Step b), and we varied guidance, that is, some learners received an explanation of the answer after being told whether they were correct (Step d) and some did not (No Step d). The scenarios for the four treatment groups in Experiment 1 are summarized in the top section of Table 2. We expect the highest levels of transfer performance for the

Table 2
Scenarios for Each Group in Experiments 1, 2, and 3

Group	Steps			
Experiment 1				
G-R (guidance–reflection)	a	b	c	d
G-NR (guidance–no reflection)	a		c	d
NG-R (no guidance–reflection)	a	b	c	
NG-NR (no guidance–no reflection)	a		c	
Experiment 2				
I-R (interactive–reflection)	a	b	c	d
I-NR (interactive–no reflection)	a		c	d
NI-R (noninteractive–reflection)		b	c	d
NI-NR (noninteractive–no reflection)			c	d
Experiment 3				
I-SR (interactive–self-reflection)	a	b	c	d
I-NR (interactive–no reflection)	a		c	d
NI-PR (noninteractive–program reflection)		b	c	d
NI-NR (noninteractive–no reflection)			c	d
I-PR (interactive–program reflection)	a	b	c	d

Note. The G-R group in Experiment 1, the I-R group in Experiment 2, and the I-SR group are identical; the G-NR group in Experiment 1 and the I-NR groups in Experiments 2 and 3 are identical; the NI-R group in Experiment 2 and the NI-PR group in Experiment 3 are identical; the NI-NR groups in Experiments 2 and 3 are identical. a = learner clicks on first answer (interactivity step); b = learner gives explanation of answer (directed reflection step); c = program tells if answer is correct (feedback step); d = program gives explanation of correct answer (guidance step).

group that received guidance, reflection, and interactivity (G-R) and the group that received guidance and interactivity with no reflection (G-NR) because guidance primes the selecting process, and reflection or interactivity primes the organizing and integrating processes. In contrast, we expect lower levels of transfer performance from the group that received no guidance along with reflection and interactivity (NG-R) or the group that received no guidance and no reflection along with interactivity (NG-NR) because the lack of guidance minimizes priming of the process of selecting—one of the key components in meaningful learning.

Method

Participants and design. The participants were 105 undergraduate freshmen from the psychology participant pool at a southwestern university (74 women and 31 men). The mean age of the participants was 18.44 ($SD = 0.75$). Each participant served in one cell of a 2 × 2 between-subjects factorial design, with the first factor being whether students were given guidance in the form of explanatory feedback (guidance and no-guidance groups, respectively) and the second factor being whether students were asked to reflect on their choices during problem solving by elaborative interrogation (reflection and no-reflection groups, respectively). There were 30 participants in the G-R group, 25 participants in the G-NR group, 26 participants in the NG-R group, and 24 participants in the NG-NR group. All participants indicated that they lacked experience in botany. Comparisons were made among the four groups on measures of retention, transfer, and program ratings.

Materials and apparatus. For each participant, the paper-and-pencil materials consisted of a participant questionnaire, a retention test, a seven-page problem-solving test, and a program-rating sheet, with each typed on an 8.5 × 11-in. sheet of paper. The participant questionnaire solicited information concerning the participant's name, gender, age, and botany knowledge. To measure the participant's knowledge of botany, we asked the following two questions: (a) "Please put a check mark indicating your knowledge of botany," followed by five blanks ranging from *very little* (scored as 0 points) to *very much* (scored as 4 points), and (b) "Please place a check mark next to the items that apply to you: ____ I have taken a class in botany. ____ I have houseplants. ____ I have eaten a plant or vegetable that I grew myself. ____ I have made my own mulch. ____ I know what a pistil is. ____ I know why plant leaves are green."

The retention test, problem-solving test, and program-rating sheet were identical to the ones used by Moreno and Mayer (2002a). The retention test included the following three questions typed on the same sheet: (a) "Please write down all the types of roots that you can remember from the lesson," (b) "Please write down all the types of stems that you can remember from the lesson," and (c) "Please write down all the types of leaves that you can remember from the lesson."

The problem-solving test consisted of seven questions. The first five questions had the following statement at the top: (a) "Design a plant to live in an environment that has low sunlight," (b) "Design a plant to live in an environment that has low temperature and high water table," (c) "Design a plant to live in an environment that has high temperature," (d) "Design a plant to live in an environment that has heavy rainfall and low nutrients," and (e) "Design a plant to live in an environment that has high wind." Students were asked to check at least one of the possible kinds of roots, stems, and leaves from a list containing all possible options and write an explanation of the choices. The last two questions had the following question and instruction at the top: "In what kind of environment would you expect to see the following plant flourish (i.e., to see the plant grow well)? Please put a check mark next to one or more conditions." A diagram of a different plant for the two problems was presented on the middle of the sheet and a list with the eight possible environmental conditions was provided under the diagram (i.e., low temperature, high temperature, low rainfall, heavy rainfall, low nutrients, high nutrients, low water table, and high water table). Additionally, at the bottom of the sheet, the following question was presented: "Why do you think that the plant designed will flourish in the environment that you chose?" Problem 6 presented a plant with thick, large, and thin-skinned leaves; short, thick, and no-bark stem; and branching, shallow, and thin roots; and Problem 7 presented a plant with thick, small, thick-skinned leaves; thick, long, and bark stem; and nonbranching, deep, and thick roots. All problem-solving questions were presented on separate sheets.

The program-rating sheet contained eight questions asking participants to rate their level of motivation, interest, understanding, and the perceived difficulty and friendliness of the program on a 10-point scale. The following two questions were intended to assess students' interest level: "How interesting is this material?" (with 1 as *boring* and 10 as *interesting*) and "How entertaining is this material?" (with 1 as *tiresome* and 10 as *entertaining*). The following question was intended to assess students' motivation level: "If you had a chance to use this program with new environmental conditions, how eager would you be to do so?" (with 1 as *not eager* and 10 as *very eager*). The following two questions were intended to assess students' level of understanding: "How much does this material help you understand the relation between plant design and the environment?" (with 1 as *not at all* and 10 as *very much*) and "How helpful is this material for learning about plant design?" (with 1 as *unhelpful* and 10 as *helpful*). The following two questions were intended to assess students' perception of learning difficulty: "How difficult was the material?" (with 1 as *easy* and 10 as *difficult*) and "How much effort is required to learn the material?" (with 1 as *little* and 10 as *much*). The following question was intended to assess students' rating of the program's friendliness: "How friendly was the computer that you interacted with?" (with 1 as *not very friendly* and 10 as *very friendly*).

The computerized materials were based on an interactive program called Design-A-Plant in which the student travels to five different alien planets with certain weather conditions, such as low rainfall or light sunlight, and must design the characteristics of the leaves, stem, and roots for a plant that would flourish there (Lester, Stone, & Stelling, 1999). It included a software pedagogical agent, Herman-the-Bug, that offered individualized advice concerning the relation between plant features and weather features by providing students with feedback on the selections they make in the process of designing plants. The feedback for each choice consisted of a verbal explanation in the form of narration. For each of the choices of roots, stem, and leaves, students were presented with the corresponding library of plant parts' graphics and names and asked to click on one of the possible options to design their plant. The G-R version of the program consisted of the seven steps (a to g) described previously. The other three versions were identical, except that the G-NR version did not include Step b, the NG-R version did not include Step d, and the NG-NR version did not include Steps b and d. The multimedia programs were developed using Director 4.04 (Macromedia, 1995a) and SoundEdit 16 (Version 2; Macromedia, 1995b). The apparatus consisted of five Macintosh IIci computer systems, which each included a 14-in. monitor, Sony headphones, and Sony audiotape recorders.

Procedure. Participants were tested in groups of 1 to 3 per session. Each participant was randomly assigned to a treatment group (G-R, G-NR, NG-R, NG-NR) and was seated at an individual cubicle in front of a computer. First, participants completed the participant questionnaire at their own rate. Second, the experimenter presented oral instructions stating that the computer program would teach them how plants should be designed in order to survive in different environments and that once the program was over, they would be tested on what they had learned. In addition, for the reflective treatment groups (G-R and NG-R), participants were informed that during their interaction they would be asked some questions, which had to be answered orally and clearly so that they could be recorded in an audiotape. Students were told to remain quietly seated once the multimedia lesson was over and that, following the presentation, they would be asked to answer a set of questions to examine how much they had learned. Participants were told to put on headphones and instructed to press the space bar to begin the program. Third, on pressing the space bar, the participants were presented once with their respective version of the multimedia program. All participants visited five different environments at their own pace. Fourth, when the program was finished, the experimenter presented oral instructions for the test, stating that there would be a series of question sheets and that for each, the participant should keep working until told to stop. Fifth, the retention sheet was distributed. After 5 min, the sheet was collected. Then, the seven problem-solving sheets were presented one at a time for 3 min each, with each sheet collected by the experimenter before the subsequent sheet was handed out. Finally, the program-rating sheet was presented and collected after 3 min. Participants were thanked for their participation and debriefed.

Scoring. A scorer not aware of the treatment condition of each participant determined the botany experience, retention, transfer, and program rating scores. The botany experience score was computed from each participant's questionnaire by adding all the check marks from the six-item botany knowledge checklist plus the participant's self-rating score (ranging from 1 point for checking *very little* to 5 points for checking *very much*). Because this questionnaire was designed to exclude those students who reported significant prior experience in the subject domain, data for students who scored above 6 were eliminated and new students were run in their places ($n = 11$).

A retention score was computed for each participant by counting the number of correct categories (out of nine possible) for each plant part (root, stem, and leaf) that the participant produced on the retention test. A close-transfer score was computed by adding up the number of right answers that students had circled on the multiple-choice portion of Transfer Problems 1 to 7. For each of the first five questions, 1 point was given by counting the number of correct categories that the participant circled for each plant part. For example, for Transfer Problem 2, which asked the student to "Design a plant to live in an environment that has low temperature and high water table," eight plant categories (branching roots, deep roots, thick roots, thick stem, bark stem, thick leaves, small leaves, thick-skinned leaves) had to be checked, and the student could obtain a maximum possible score of 8 points on that question. For each of the last two questions, 1 point was given for each correct environment condition chosen by the participant (with a maximum possible of 4). In addition, we computed a far-transfer score for each participant by counting the number of acceptable explanations that the participant produced across the seven transfer problems. For each of the first five problems, 1 point was given for each correct explanation corresponding to the plant categories checked by the student. For each of the last two problems, 1 point was given for each correctly stated explanation about the participant's choice of type of environment, regardless of wording. A program-ratings score was computed by adding up the mean number that students had circled on the five program ratings (motivation, interest, understanding, friendliness, and perceived difficulty).

For students in the reflective conditions (G-R and NG-R groups), the pedagogical agent asked them to give oral explanations (Step b) for each of their plant design choices (Step a). Students' protocols were transcribed, and their answers were scored as correct, appropriate theory, or not correct, misconception or missing theory. To classify students' explanations as correct or incorrect, we compared the explanation for designing each plant part with the explanation that had been explicitly provided by the guided version of the multimedia program. Answers that corresponded to the guided explanation were classified as correct. For example, for the problem of choosing the appropriate roots for a plant that needs to survive in an environment with low rain, a correct answer consisted of explaining that shallow and branching roots survive best in low rain because they are able to spread and absorb the scarce water that falls on the surface of the ground. For that same problem, an example of an incorrect explanation consisted of stating that "Deep roots survive best in low rain because they can reach deep into the soil to look for water." Similarly, answering "I do not know," "I was just guessing," or "It looks like it would survive" were scored as incorrect explanations. From these data, we obtained two scores for each student. First, we added the total number of wrong explanations that students had given during Step b to the plant designed during Step a. Second, we computed the proportion of learned answers for each student by counting the number of times that students had changed their wrong designs (Step a) to correct designs (Step f) after the agent's corrective or explanatory feedback (for no-guidance and guidance groups, respectively) and dividing this number by the total number of wrong designs given during Step a.

Results and Discussion

Table 3 shows the mean scores and standard deviations for the G-R, G-NR, NG-R, and NG-NR groups on measures of retention, close and far transfer, and program ratings (which had an internal consistency reliability of $\alpha = .84$). A two-factor multivariate analysis of variance (MANOVA) was conducted, with guidance (guidance vs. no guidance) and reflection (reflection vs. no reflection) as between-subjects factors, and retention, close and far transfer, and program ratings as dependent measures.

The assumption of homogeneity of variance was tested prior to the MANOVA and found to be tenable, Box's M (30, 26665) = 43.81, $p = .09$. Significant differences were found among the guidance and nonguidance groups on the dependent measures (Wilks's $\Lambda = .64$), $F(4, 98) = 14.01$, $p < .01$. Conversely, no

Table 3
Mean Score on Retention and Transfer Tests and Program Ratings and Corresponding Standard Deviations for Four Groups in Experiment 1

	Type of test							
	Retention		Close transfer		Far transfer		Program ratings	
Group	*M*	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*
G-R	6.83	1.60	21.37	3.81	12.93	5.02	32.20	6.35
G-NR	7.08	1.75	19.84	3.58	12.96	4.95	32.04	5.63
NG-R	6.58	1.36	18.04	3.79	6.38	3.50	28.50	6.68
NG-NR	6.33	1.47	17.92	3.27	6.67	3.45	30.46	6.65

Note. Potential scores ranged from 0 to 9 for the retention test, from 0 to 32 for the close-transfer test, from 0 to 28 for the far-transfer test, and from 5 to 50 for the program-rating scores. G-R = guidance–reflection; G-NR = guidance–no reflection; NG-R = no guidance–reflection; NG-NR = no guidance–no reflection.

significant differences were found among the reflective and nonreflective groups on the dependent measures (Wilks's $\Lambda = .97$), $F(4, 98) = 0.66$, $p = .62$. Finally, there were no significant differences in the interactions between guidance and reflection on the dependent measures (Wilks's Λ) = .97, $F(4, 98) = 0.90$, $p = .47$. Two-way analyses of variances (ANOVAs) on each dependent variable were conducted as follow-up tests to the MANOVA. Using the Bonferroni method to adjust for Type I error, we tested each ANOVA at an alpha level of .0125.

Issue 1: Do students who are guided in the design of plant parts learn better from an agent-based multimedia program than students who learn by discovery? According to the cognitive model of multimedia learning, guidance is expected to promote meaningful learning because it encourages the learner to engage in the cognitive process of selecting—a basic component process in meaningful learning that might otherwise not be primed. Thus, one of the hypotheses that our study was designed to examine was that scaffolding students' learning by the use of guidance would be more likely to promote students' understanding of an interactive multimedia lesson than a lesson that did not include guidance. The two-factor ANOVA failed to revealed a main effect for guidance on retention, $F(1, 101) = 4.68$, $MSE = 10.07$, $p = .03$. Groups that were presented with guided explanations only marginally recalled more information about the plant library than those that were not presented with the agent's explanations (*M*s = 7.05 and 6.46; *SD*s = 1.52 and 1.40, respectively), yielding an effect size of 0.42.

There was a main effect for guidance on both transfer measures, $F(1, 101) = 13.62$ and 57.24, $MSE = 179.71$ and 1,074.66, $p < .01$, for close- and far-transfer scores, respectively. The mean number of correct close-transfer answers was 20.60 and 17.98, respectively, for the guidance and no-guidance groups (*SD*s = 3.75 and 3.51, respectively). The mean number of correct far-transfer answers was 12.95 and 6.53, respectively, for the guidance and no-guidance groups (*SD*s = 4.94 and 3.44, respectively). Groups presented with the agent's explanatory feedback gave significantly more correct answers on the transfer tests than those presented solely with information on the correctness of their answers, yielding an effect size of 0.75 and 1.87 for the close- and far-transfer tests, respectively. Finally, the groups differed marginally on program ratings, $F(1, 101) = 4.52$, $MSE = 181.78$, $p = .04$, with a mean rating of 32.13 and 29.44, respectively, for the guidance and no-guidance groups (*SD*s = 5.98 and 6.67, respectively). The effect size was 0.40.

Issue 2: How does guidance help students' learning? On the basis of a cognitive model of multimedia learning, we expected that students in the G-R group would give more correct explanations during reflection and show greater error reduction in their plant designs as compared with students in the NG-R group. To test this thesis, we recorded the number of times that each student gave a wrong explanation during reflection due to misconceptions or lacking explanations for their plant design choices (Step b). The protocols for 3 students were lost because of a malfunction of one of the audiotape recorders. Using the number of wrong answers given during reflection as a dependent measure, we performed a one-factor ANOVA that revealed a main treatment effect, $F(1, 51) = 16.57$, $MSE = 91.00$, $p < .01$. Students in the G-R group gave significantly fewer wrong answers than did students in the NG-R group during their interaction with the program. The mean number of wrong answers was 6.15 and 8.77 (*SD*s = 1.81 and 2.79) for the G-R and NG-R groups, respectively. The effect size was 0.94.

In addition, we conducted a one-factor ANOVA using group (G-R vs. NG-R) as the between-subjects factor and the proportion of learned answers as the dependent measure. The ANOVA provided evidence for a significant treatment effect, $F(1, 51) = 48.90$, $MSE = 2.86$, $p < .01$. The mean proportion of learned answers was .82 and .35 (*SD*s = .23 and .25) for the G-R and NG-R groups, respectively. The effect size was 1.88.

In sum, these findings demonstrate a *guidance effect* in agent-based multimedia. Students achieve better transfer scores, give more correct explanations for their choices, and change their wrong answers to right answers significantly more when the agent provides guidance in the form of explanatory feedback rather than when the agent provides corrective feedback alone.

Issue 3: Do students who are asked to give explanations about their answers learn better from an agent-based multimedia program than students who are not asked to give explanations? According to the cognitive theory of multimedia learning, reflection should not have a strong effect on meaningful learning because the interactivity in the game already primes the cognitive processes of organizing and integrating. The two-factor ANOVA failed to reveal a main effect for reflection on retention, $F(1, 101) = 0.18$, $MSE = 0.38$, $p = .67$. Groups that were asked to explain why they had chosen the respective plant parts did not differ in their recall of general information about the plant library from those that were not (*M*s = 6.71 and 6.84; *SD*s = 1.49 and 1.50, respectively). There was no significant interaction between guidance and reflection, $F(1, 101) = 1.61$, $MSE = 3.47$, $p = .21$.

The ANOVAs failed to reveal a main effect for reflection on both transfer measures, $F(1, 101) = 1.34$ and 0.03, $MSE = 17.71$ and 0.62, $p = .25$ and 0.86, for close and far transfer, respectively. Reflective groups did not differ in their performance on close-transfer tests (*M*s = 19.82 and 18.90; *SD*s = 4.12 and 3.53, respectively) nor did they differ in their performance on far-transfer tests (*M*s = 9.89 and 9.88; *SD*s = 5.45 and 5.29, respectively) from nonreflective groups. There was no significant interaction between guidance and reflection, $F(1, 101) = 0.97$ and 0.02, $MSE = 12.86$ and 0.43, $p = .33$ and .88, for close and far transfer, respectively.

Finally, there was no main effect for reflection on program ratings, $F(1, 101) = 0.52$, $MSE = 21.07$, $p = .47$. The mean respective ratings for the reflective and no-reflective groups were 30.48 and 31.26 (*SDs* = 6.71 and 6.14), respectively. There was no significant interaction between guidance and reflection, $F(1, 101) = 0.73$, $MSE = 29.24$, $p = .40$. Overall, no evidence was found for a reflection effect. Asking students to give explanations about their problem-solving choices did not affect their learning or their impressions about the agent-based multimedia game.

Experiment 2

Experiment 1 demonstrated the beneficial effects of designing agent-based interactive multimedia games with guidance in the form of explanatory feedback for low-experience learners. However, contrary to past research in reading, including a reflection technique did not help students process the materials more deeply. Our interpretation for the different pattern is that the effects of reflection treatments are diminished in the context of an interactive multimedia learning task. When students are asked to make choices during a learning task, such as when they need to click on a choice among a set of possible alternatives during problem solving, their cognitive activity—that is, organizing and integrating processes—is already at a high level. Students who learn in the botany game need to make inferences between the plant structure and the weather conditions when deciding on a particular plant design. Thus, asking students to explain why they made a certain decision is not as crucial as asking students "why" questions after reading text. Because reading is an automatic process, it can be accomplished without encoding meaning. Therefore, an elaborative interrogation technique during reading is an effective technique to encourage students to use their cognitive resources to reflect on the content of a lesson (King, 1992; Pressley et al., 1992; Woloshyn et al., 1994).

If this interpretation is correct, using reflection techniques in noninteractive multimedia programs where students are not asked to make choices (noninteractive games) should have a similar advantage to that observed during reading (a noninteractive learning task). In noninteractive or linear multimedia environments, meaning is constructed from mentally selecting, organizing, and integrating pictures and words into a meaningful model but without engaging in behavioral decisions during the lesson. Experiment 2 was designed to test this hypothesis by comparing the learning outcomes of students who were presented with a reflection technique in either interactive or noninteractive conditions.

In Experiment 2, all learners were told the correct answer (Step c) and were given an explanation for the correct answer (Step d). We varied interactivity, that is, some learners were asked to click on an answer (Step a) and some were not (No Step a), and we varied reflection, that is, some learners were asked to give an explanation for the answer (Step b) and some were not (No Step b). The scenarios for the four treatment groups in Experiment 2 are summarized in the middle section of Table 2. As you can see, students in the interactive and reflection (I-R) group provided an explanation for their answer (which could be right or wrong), whereas students in the noninteractive and reflection (NI-R) group provided an explanation for the answer given by the agent (which was right).

According to the cognitive theory of multimedia learning, the group that receives guidance and reflection on the right answer (NI-R) should perform best on transfer tests because guidance primes the selecting process and reflection on the correct answer primes the processes of organizing and integrating. This arrangement is particularly powerful because the learner's deep cognitive processing (i.e., organizing and integrating) is more efficiently coordinated with the correct material (i.e., selecting).

Method

Participants and design. The participants were 71 undergraduate freshmen from the psychology participant pool at a southwestern university (51 women and 20 men). The mean age of the participants was 20.44 (SD = 4.54). Each participant served in one cell of a 2 × 2 between-subjects factorial design, with the first factor being whether students were able to interact during learning (interactive and noninteractive groups, respectively) and the second factor being whether students were asked to reflect on their choices during problem solving by elaborative interrogation (reflection and no-reflection groups, respectively). There were 19 participants in the I-R group, 17 participants in the interactive and no-reflection (I-NR) group, 17 participants in the NI-R group, and 18 participants in the noninteractive and no-reflection (NI-NR) group. All participants scored low (i.e., 6 or below) on an 11-point scale of botany knowledge. In a similar way to Experiment 1, comparisons were made among the four groups on measures of retention, close and far transfer, and program ratings.

Materials and apparatus. For each participant, the paper-and-pencil materials and apparatus were identical to those used in Experiment 1. They consisted of a participant questionnaire, a retention test, a seven-page problem-solving test, and a program-rating sheet, with each typed on 8.5 × 11-in. sheets of paper. The computerized materials consisted of four multimedia computer programs on how to design a plant (Lester et al., 1999). The I-R version was identical to the G-R version for Experiment 1, and the I-NR version was identical to the G-NR version for Experiment 1. The noninteractive versions of Experiment 2 (NI-R and NI-NR) were identical to the corresponding interactive versions (I-R and I-NR), with the exception that for the same set of five different environmental conditions, the pedagogical agent does not ask the student to design the plant part that is appropriate for that environment. In addition, the NI-R condition differed from the I-R condition in that the agent asks students to give explanations for the program choices rather than for students' choices by asking: "Why do you think that particular type of root/stem/leaf will survive in this environment?" The multimedia programs were developed using Director 4.04 (Macromedia, 1995a) and SoundEdit 16 (Version 2; Macromedia, 1995b). The apparatus consisted of six Pentium III PC systems, which each included a 15-in. monitor, Sony headphones, and Sony audiotape recorders.

Procedure. The procedure was identical to that used in Experiment 1, except that each participant was assigned to either the I-R, I-NR, NI-R, or NI-NR group. Additionally, we did not analyze students' protocols for the reflection treatments because one of the conditions did not include students' interactions (the NI-R group), so looking at the number of times that students had changed their wrong plant designs to correct plant designs after elaborative interrogation was not possible. Despite this fact, for consistency purposes, we told students that their responses were tape-recorded for further analysis.

Scoring. The botany experience score, retention score, transfer scores, and program-rating score for each participant were computed identically to the way they were computed in Experiment 1. Data for students who scored above 6 in the botany experience questionnaire were eliminated and new students were run in their places (n = 8).

Results and Discussion

Table 4 shows the mean scores and standard deviations for the I-R, I-NR, NI-R, and NI-NR groups on measures of retention, close and far transfer, and program ratings (which had an internal consistency reliability of α = .86). A two-factor MANOVA was conducted, with interactivity (interactive vs. noninteractive) and reflection (reflection via elaborative interrogation vs. no reflection) as the between-subjects factors, and retention, close and far transfer, and program ratings as dependent measures.

We conducted tests of homogeneity of variance prior to the MANOVA and found that the assumption was met, Box's M(30, 12139) = 45.92, p = .09. Significant differences were found among the interactive and noninteractive groups on the dependent measures (Wilks's Λ = .79), $F(4, 64) = 4.25$, $p < .01$. In addition, significant differences were found among the reflective and non-reflective groups on the dependent measures (Wilks's Λ = .84), $F(4, 64) = 3.10$, $p = .02$. Finally, there were significant differences in the interactions between interactivity and reflective methods on the dependent measures (Wilks's Λ = .81), $F(4, 64) = 3.76$, $p < .01$. Two-way ANOVAs on each dependent variable

Table 4
Mean Score on Retention and Transfer Tests and Program Ratings and Corresponding Standard Deviations for Four Groups in Experiment 2

	Type of test							
	Retention		Close transfer		Far transfer		Program ratings	
Group	*M*	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*
I-R	6.68	1.63	21.26	3.16	12.10	4.69	32.79	5.71
I-NR	7.59	1.23	19.82	4.10	12.71	4.93	32.68	4.69
NI-R	7.47	1.37	24.76	5.06	20.06	6.51	34.85	7.66
NI-NR	6.33	1.71	20.11	3.80	13.56	6.64	30.67	7.91

Note. Potential scores ranged from 0 to 9 for the retention test, from 0 to 32 for the close-transfer test, from 0 to 28 for the far-transfer test, and from 5 to 50 for the program-rating scores. I-R = interactive–reflection; I-NR = interactive–no reflection; NI-R = noninteractive–reflection; NI-NR = noninteractive–no reflection.

were conducted as follow-up tests to the MANOVA. Using the Bonferroni method to adjust for Type I error, we tested each ANOVA at an alpha level of .0125.

Issue 1: Do students who interact in the design of plant parts learn better from an agent-based multimedia program than students who learn without interacting? The first issue that Experiment 2 was designed to examine was whether having students behaviorally interact by having to choose among a set of alternatives would be more likely to promote students' understanding of a multimedia lesson than a lesson that did not include students' interaction. The two-factor ANOVA failed to reveal a main effect for interactivity on retention, $F(1, 67) = 0.43$, $MSE = 0.97$, $p = .52$ (*M*s = 7.11 and 6.89; *SD*s = 1.51 and 1.64, respectively, for interactive and noninteractive groups). Groups differed only marginally on their mean close-transfer scores, $F(1, 67) = 3.86$, $MSE = 63.58$, $p = .05$ (*M*s = 20.58 and 22.37; *SD*s = 3.65 and 4.98, respectively, for interactive and noninteractive groups). In addition, the two-factor ANOVA failed to reveal a main effect for interactivity on program ratings, $F(1, 67) = 0.00$, $MSE = .00$, $p = .99$, with a mean ratings score of 32.74 and 32.70 (*SD*s = 5.18 and 7.97), respectively, for the interactive and noninteractive groups. Conversely, there was a main effect for interactivity on far-transfer scores, $F(1, 67) = 14.77$, $MSE = 629.26$, $p < .01$, with a mean rating of 12.39 and 18.23 (*SD*s = 4.75 and 9.20), respectively, for the interactive and noninteractive groups. The effect size was 0.63.

Issue 2: Do students who are asked to give explanations learn better from an agent-based multimedia program than students who are not asked to give explanations? The main issue to be examined in Experiment 2 was whether asking students to explain their answers (I-R group) or the answers provided by the program (NI-R group) with an elaborative interrogation technique would be more likely to promote students' understanding of the scientific system than presenting the same materials without asking students to give explanations.

The two-factor ANOVA failed to reveal a main effect for reflection on retention, $F(1, 67) = 0.11$, $MSE = 0.24$, $p = .75$ (*M*s = 7.06 and 6.94; *SD*s = 1.55 and 1.61, respectively, for R and NR groups). In addition, there was no main effect on program ratings, $F(1, 67) = 1.86$, $MSE = 81.85$, $p = .18$. The mean ratings for the reflection and no-reflection groups were 33.76 and 31.64 (*SDs* = 6.69 and 6.54), respectively.

On the other hand, a significant main effect was found on both transfer measures, $F(1, 67) = 9.99$ and 8.46, $MSE = 164.40$ and 360.29, $ps < .01$, for close and far transfer, respectively, yielding an effect size of 0.76 and 0.72, respectively. The mean number of correct answers on the close-transfer tests was 22.92 and 19.97 (*SD*s = 4.47 and 3.89), respectively, for the reflective and no-reflective groups. The mean number of correct answers on the far-transfer tests was 17.33 and 13.14 (*SD*s = 8.97 and 5.81), respectively, for the reflective and no-reflective groups.

Despite the main effect on reflection for transfer measures, the analysis that is most relevant to the hypothesis raised in Experiment 2 is the interaction between reflection and interactivity. No interaction was found between interactivity and reflection for program ratings or close-transfer tests, $F(1, 67) = 1.67$ and 2.78, $MSE = 73.47$ and 45.74, $p = .20$ and .10, respectively. Conversely, there was a significant interaction between interactivity and reflection for retention and far-transfer measures, which was important to our predictions, $F(1, 67) = 8.11$ and 10.86, $MSE = 18.45$ and 462.64, $ps < .01$, respectively. Within the interactions, we examined the interactivity and reflection simple main effects with supplemental post hoc Tukey's tests (at $\alpha = .05$).

Consistent with our predictions, for noninteractive learning conditions, students who learned with reflective techniques (NI-R group) remembered significantly more of the plant library, $F(1, 67) = 4.97$, $MSE = 11.31$, $p < .05$, and had higher far-transfer scores, $F(1, 67) = 19.00$, $MSE = 809.26$, $p < .01$, than those who learned with nonreflective techniques (NI-NR group). Conversely, we found that for interactive learning conditions (I-R and I-NR groups), there were no significant differences on retention or far-transfer scores, $F(1, 67) = 3.22$ and 0.08, $MSE = 7.33$ and 3.24, $p = .08$ and .78, respectively.

For nonreflective learning conditions, students who learned with interactive programs (I-NR group) remembered significantly more of the plant library, $F(1, 67) = 6.05$, $MSE = 13.77$, $p = .02$, but did not show higher far-transfer scores than students who learned with noninteractive programs (NI-NR group), $F(1, 67) = 0.15$, $MSE = 6.31$, $p = .70$. For reflective learning conditions, students who learned with interactive and noninteractive programs (I-R and NI-R groups) did not differ on measures of retention, $F(1, 67) = 2.44$, $MSE = 5.55$, $p = .12$. However, the NI-R group outperformed the I-R group on far-transfer scores, $F(1, 67) = 25.82$, $MSE = 1{,}099.74$, $p < .01$.

In sum, interactivity helped students' retention when learning from nonreflective versions of the program, and reflection helped students' retention and far transfer when learning from noninteractive versions of the program. As can be seen from looking at the first two rows of Tables 3 and 4, the results from Experiments 1 and 2 show that reflective and nonreflective interactive groups (G-R and G-NR for Experiment 1 and I-R and I-NR for Experiment 2) did not differ on any of the dependent measures used in the studies. Taken together, these results support the hypothesis that reflection techniques help students learn from noninteractive conditions but not from interactive environments. However, further research is needed to examine the main effects found in this experiment.

A seemingly contradictory finding was that students who learned with no interactivity and reflection (NI-R group) outperformed those who learned with interactivity and reflection (I-R group) on tests of far problem-solving transfer. A possible explanation for this result can be offered by taking into consideration what kind of information students were asked to elaborate on. Whereas for the I-R group, the agent asked students to explain in words why they selected a particular plant type, for the NI-R group, the agent asked students to explain in words why the program selected a particular plant type. Similar to other cases of elaborative interrogation that have been found to have positive effects on students' retention and comprehension (Bruning et al., 1999), students in the NI-R group were asked to reflect on correct information that they had just been presented with. On the other hand, students in the I-R group were asked to reflect on information that they had generated and therefore may not have been correct. In sum, our explanation for this puzzling effect is that using elaborative interrogation before students get corrective feedback may promote the consolidation of an incorrect mental model by having students verbalize their misconceptions. Although both NI-R and I-R groups were presented with an identical explanation (Step d) and correct example (Step g), it is likely that verbalizing misconceptions may counteract the program's feedback for the I-R group. Experiment 3 was designed to test this hypothesis.

Experiment 3

The first goal of Experiment 3 was to determine whether it would be possible to replicate the findings of Experiment 2. The second goal was to determine more precisely the cause of far-transfer differences between the interactive and noninteractive groups that learned with reflection in Experiment 2 (assuming that the difference could be replicated). Experiment 3 was identical to Experiment 2 in that we compared the learning outcomes of students who were presented with a reflection technique in either interactive or noninteractive conditions. However, to test the hypothesis that the beneficial effects of reflection are contingent on the quality of the elaboration made by students, we added an interactive condition where students were asked to reflect on the program's correct solutions rather than on their own.

Experiment 3 consisted of the same four groups as in Experiment 2, along with a fifth group, interactive and program reflection (I-PR). In Experiment 3, all learners were told the correct answer (Step c) and were given an explanation for the correct answer (Step d). As in Experiment 2, we varied interactivity, that is, some learners were asked to click on an answer (Step a) and some were not (No Step a), and we varied reflection, that is, some learners were asked to give an explanation for the answer (Step b) and some were not (No Step b). As in Experiment 2, in the group that received interactivity and self-reflection (I-SR), learners generated an explanation for their answer (which might be right or wrong), and in the group that received no interactivity and program reflection (NI-PR), learners generated an explanation for the correct answer given by the on-screen agent. In addition, we added a fifth group (I-PR) that included interactivity (Step a) and reflection (Step b) but with the reflection occurring after receiving the correct answer (in Step c) so the learner's explanation was for the correct answer. The scenarios for the five treatment groups in Experiment 3 are summarized in the bottom section of Table 2. On the basis of the cognitive theory of multimedia learning, we expect the best transfer performance from the two groups that reflected on the correct material—the NI-PR group and the I-PR group—because both engaged in deep processing (i.e., organizing and integration) coordinated with a focus on the relevant material (i.e., selecting).

Method

Participants and design. The participants were 78 undergraduate freshmen from the psychology participant pool at a southwestern university (54 women and 24 men). The mean age of the participants was 19.72 ($SD = 2.19$). There were 15 participants in the I-NR group, 16 participants in the I-SR group, 15 participants in the I-PR group, 16 participants in the NI-PR group, and 16 participants in the NI-NR group. All participants scored low (i.e., 6 or below) on an 11-point scale of botany knowledge. Comparisons were made among the five groups on measures of retention, close and far transfer, and program ratings. In addition, comparisons were made among the three reflective groups (I-SR, I-PR, and NI-PR) on the quality of the explanations that were verbalized during elaborative interrogation (no explanation, wrong explanation or misconception, right explanation).

Materials and apparatus. For each participant, the paper-and-pencil materials and apparatus were identical to those used in Experiments 1 and 2. They consisted of a participant questionnaire, a retention test, a seven-page problem-solving test, and a program-rating sheet, with each typed on 8.5×11-in. sheets of paper. The computerized materials and apparatus for the I-NR, I-SR, NI-PR, and NI-NR conditions were identical to those used in the I-NR, I-R, NI-R, and NI-NR conditions for Experiment 2, respectively. The I-PR version was identical to the I-SR version with one exception: rather than having students reflect on the design of their own plant, they were asked to elaborate on the correct plant design presented by the multimedia program.

Procedure. The procedure was identical to that used in Experiment 2 except that we recorded, transcribed, and analyzed students' protocols for the reflective treatments (I-SR, I-PR, and NI-PR groups).

Scoring. The botany knowledge, retention, close- and far-transfer, and program-rating scores for each participant were computed identically to the way they were computed in Experiments 1 and 2. Data for students who scored above 6 in the botany experience questionnaire were eliminated and new students were run in their places ($n = 7$). In addition, for students in the reflective conditions (I-SR, I-PR, and NI-PR groups), we transcribed and scored their explanations during reflective elaboration as correct, incorrect, or missing theory. Similar to Experiment 1, to classify students' explanations as correct or incorrect, we compared the explanation for designing each plant part with the explanation that had been explicitly provided by the multimedia program. Answers that corresponded to the program's explanation were classified as correct and those that did not correspond to the program's explanations were classified as incorrect.

Finally, when students failed to provide an answer or stated that they did not know the answer, we classified the response as missing. From these data, we obtained three scores indicating the proportion of correct, incorrect, and missing explanations for each student by counting the total number of correct, incorrect, and missing explanations, respectively, and dividing those totals by the overall number of explanations given during elaborative interrogation.

Results and Discussion

Table 5 shows the mean scores and standard deviations for the I-SR, I-PR, I-NR, NI-PR, and NI-NR groups on measures of retention, close transfer, far transfer, and program ratings (which had an internal consistency reliability of $\alpha = .88$). We conducted a MANOVA, with retention, close and far transfer, and program ratings as dependent measures. Tests of homogeneity of variance revealed that the assumption was met, Box's $M(40, 11665) = 68.13$, $p = .06$. Significant differences were found among the groups on the dependent measures (Wilks's $\Lambda = .63$), $F(16, 214) = 2.22$, $p < .01$. One-way ANOVAs on each dependent variable were conducted as follow-up tests to the MANOVA. Using the Bonferroni method to adjust for Type I error, we tested each ANOVA at an alpha level of .0125.

Issue 1: Do agent-based multimedia games that include interactivity and reflection help students learn better than those that do not? Using retention as a dependent variable, we found that the ANOVA revealed a main effect on retention, $F(4, 73) = 3.66$, $MSE = 9.23$, $p < .01$. Post hoc Tukey's tests (at $\alpha = .05$) indicated that students in reflective groups (I-PR, NI-PR, and I-SR) recalled significantly more information about the plant library than those in no-reflective groups (I-NR and NI-NR), yielding an effect size of 0.81. Groups did not differ on close transfer, $F(4, 73) = 0.97$, $MSE = 14.79$, $p = .43$. Conversely, there was a main treatment effect on far-transfer scores, $F(4, 73) = 5.61$, $MSE = 199.00$, $p < .01$. Post hoc Tukey's tests (at $\alpha = .05$) indicated that groups that were asked to reflect on the program's solutions (I-PR and NI-PR) gave significantly more creative solutions to novel problems than the rest of the groups in the far-transfer test. The effect size was 0.80. Finally, no treatment effect was found on the program ratings, $F(4, 73) = 1.03$, $MSE = 54.31$, $p = .40$. In sum, no treatment effects were obtained on close transfer and program ratings, but, similar to Experiment 2, when students were asked to reflect on correct problem-solving answers (i.e., like the NI-R group in Experiment 2), they outperformed the rest of the groups on retention and far-transfer measures. A possible interpretation for the different pattern of results for the close transfer versus retention and far-transfer measures is that the format of the close-transfer assessment (a multiple-choice graphic questionnaire) may not have been a sensitive measure of students' learning. Both retention and far-transfer measures were open-ended questions requiring students to produce rather than choose an answer.

Issue 2: Does the effect of reflection depend on the information that students are asked to reflect on? To be able to support our hypothesis that the far-transfer superiority of the NI-R group over the I-R group found in Experiment 2 may have relied on students' reflecting on correct rather than incorrect answers, we compared the proportion of correct and incorrect verbalizations for students in the I-SR, I-PR, and NI-PR groups. To do so, we conducted a MANOVA using treatment as between-participants factor and the proportion of correct and incorrect explanations that students had given during elaborative interrogation as dependent variables. The assumption of homogeneity of variance was tested prior to the MANOVA and found to be tenable, Box's $M(6, 46945) = 6.77$, $p = .39$. Significant differences were found among the treatment groups on the dependent measures (Wilks's Λ) = .68, $F(4, 86) = 4.66$, $p < .01$. One-way ANOVAs on each dependent variable were conducted as follow-up tests. Using the Bonferroni method to adjust for Type I error, we tested each ANOVA at an alpha level of .025.

Using the proportion of correct answers as a dependent variable, we found that the ANOVA revealed a main treatment effect, $F(2, 44) = 4.17$, $MSE = 0.25$, $p = .022$. Post hoc Tukey's tests (at $\alpha = .05$) indicated that students who were asked to elaborate on the program solutions (I-PR and NI-PR groups) had a higher proportion of correct explanations than those who were asked to elaborate on their own solutions to the problems. The effect size was 1.10. The mean proportion of correct explanations was 0.41, 0.62, and 0.64 (*SD*s = 0.20, 0.21, and 0.31), respectively, for the I-SR, I-PR, and NI-PR groups. In addition, there was a treatment effect on the proportion of wrong explanations offered during elaborative interrogation, $F(2, 44) = 9.03$, $MSE = 0.16$, $p < .01$. Post hoc Tukey's

Table 5
Mean Score on Retention and Transfer Tests and Program Rating and Corresponding Standard Deviations for Four Groups in Experiment 3

	Type of test							
	Retention		Close transfer		Far transfer		Program ratings	
Group	*M*	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*
I-PR	7.33	1.76	23.07	2.52	19.60	5.90	30.87	7.29
I-SR	7.19	1.42	21.19	2.95	13.44	5.73	31.16	7.00
I-NR	6.00	1.77	22.20	3.32	13.73	6.03	34.67	5.88
NI-PR	7.38	1.31	22.50	4.35	18.93	6.57	30.06	5.69
NI-NR	5.81	1.64	20.69	5.55	11.63	5.50	30.19	9.66

Note. Potential scores ranged from 0 to 9 for the retention test, from 0 to 32 for the close-transfer test, from 0 to 39 for the far-transfer test, and from 5 to 50 for the program-rating score. I-PR = interactive–program reflection; I-SR = interactive–self-reflection; I-NR = interactive–no reflection; NI-PR = noninteractive–program reflection; NI-NR = noninteractive–no reflection.

tests (at $\alpha = .05$) indicated that the NI-PR group gave a significantly lower proportion of incorrect explanations than the I-PR and I-SR groups and the I-PR group gave a significantly lower proportion of incorrect explanations than the I-SR group, yielding an effect size of 1.14. The mean proportion of incorrect explanations was 0.38, 0.27, and 0.17 (*SD*s = 0.14, 0.15, and 0.12), respectively, for the I-SR, I-PR, and NI-PR groups. In sum, consistent with the hypothesis raised in Experiment 3, for reflection to be effective, students must be asked to reflect on correct models of the new information. Novice students who are asked to give explanations about their own models during problem solving (I-SR group) may be hurt by consolidating an incorrect model for the scientific system to be learned.

General Discussion

Our findings have useful theoretical and practical implications. On the theoretical side, the guidance effects on learning that were found in Experiment 1 are consistent with the cognitive model of multimedia learning on which they are based and consistent with a growing body of research showing that students learn more deeply from guided discovery than pure discovery (Chall, 2000; Mayer, 2002, 2004; McKeough, Lupart, & Marini, 1995; Schauble, 1990; Singley & Anderson, 1989). It is important to note that these findings show that the benefits of guidance extend into the realm of computer games and simulations.

On the other hand, the fact that Experiment 1 failed to show a reflection effect led us to distinguish between behaviorally interactive and noninteractive instructional materials. We hypothesized that when students are required to make decisions during the process of knowledge construction (such as in the case of having to decide which plant design is the most appropriate to click on), they are encouraged to engage in active cognitive processing. Adding reflection to an interactive environment does not significantly improve their learning, presumably because interactivity already primes the cognitive processes of organizing and integrating. Experiment 2 was designed to test this hypothesis.

By finding that students who learn from noninteractive multimedia significantly increase their retention and far transfer with reflection techniques, our second study seemed to reconcile the tension between the results found in Experiment 1 and past research on elaborative interrogation in the reading comprehension literature. However, a seemingly contradictory finding was that the far-transfer scores of the group of students who learned with reflection and no interactivity were significantly superior to the group of students who learned with reflection and interactivity. Experiment 3 helped clarify the relationship between reflection and interactivity by distinguishing between instructional methods that ask students to reflect on correct information (I-PR and NI-PR groups) and those that ask students to reflect on their own answers, which may be incorrect (I-SR group). The results of our third study replicated those found in our second study and gave additional evidence in support of the hypothesis that reflection alone does not foster deeper learning unless it is based on correct information. Taken together, the contribution of Experiments 2 and 3 is to help understand when the reflection techniques promote multimedia learning and why (namely, by fostering deeper cognitive processing of correct information).

On the practical side, our results have direct implications for the design of agent-based multimedia. Although pedagogical agents are widely used in instructional design, there is still a need for empirically based principles for the design of agent-based environments in educational technology (Moreno et al., 2001). A contribution of the present set of studies is to point out two possible roles that pedagogical agents may have in multimedia learning. First, designers of agent-based games should incorporate structured guidance rather than rely solely on pure discovery. Apparently, unstructured activity—such as our corrective-feedback-alone treatment—is not as effective as more direct instruction—such as our explanatory-feedback treatment. Second, an additional cognitive role that pedagogical agents may play in learning environments that lack interactivity is to promote students' reflection via elaborative interrogation techniques for correct answers. It is worthwhile to note that guidance and reflection produced substantial effect sizes under certain circumstances indicating that the effects have practical significance as well as statistical significance.

Finally, it is important to note that our research is limited because it deals with only one kind of computer game (i.e., Design-a-Plant), one kind of interactivity (i.e., having students select appropriate visual information), one kind of guidance (i.e., scientific explanations given as feedback), one kind of reflective technique (i.e., a particular kind of elaborative interrogation), and one kind of learner (i.e., college students who were unfamiliar with botany). Future research is needed to determine how to incorporate interactivity, structured guidance, and reflective thought using other educational games, methods, and learners. For example, Mayer, Mautone, and Prothero (2002) found that students learned more deeply from a computer-based geology game when they received explicit guidance concerning how to visualize relevant geological structures. Agent-based computer games and simulations offer a potentially valuable venue for science education, but inexperienced learners may need structured guidance in combination with reflective techniques to help them achieve deep understanding.

References

Anderson, R. C., & Pearson, P. D. (1984). A schema-theoretic view of basic processes in reading comprehension. In P. D. Pearson, R. Barr, M. L. Kamil, & P. Mosenthal (Eds.), *Handbook of reading research* (Vol. 1, pp. 255–291). White Plains, NY: Longman.

Brown, A., & Campione, J. (1994). Guided discovery in a community of learners. In K. McGilly (Ed.), *Classroom lessons: Integrating cognitive theory and classroom practice* (pp. 229–270). Cambridge, MA: MIT Press.

Bruner, J. S. (1961). The art of discovery. *Harvard Educational Review, 31,* 21–32.

Bruning, R. H., Schraw, G. J., & Ronning, R. R. (1999). *Cognitive psychology and instruction.* Upper Saddle River, NJ: Prentice Hall.

Chall, J. S. (2000). *The academic achievement challenge.* New York: Guilford Press.

Chi, M. T. H., de Leeuw, N., Chiu, M. H., & La Vancher, C. (1994). Eliciting self-explanations improves understanding. *Cognitive Science, 18,* 439–477.

de Jong, T., & van Joolingen, W. R. (1998). Scientific discovery learning with computer simulations of conceptual domains. *Review of Educational Research, 68,* 179–201.

Doctorow, M., Wittrock, M. C., & Marks, C. (1978). Generative processes in reading comprehension. *Journal of Educational Psychology, 70,* 109–118.

Gagne, R. M. (1965). *The conditions of learning.* New York: Holt.

Hardiman, P., Pollatsek, A., & Weil, A. (1986). Learning to understand the balance beam. *Cognition and Instruction, 3,* 1–30.

Jonassen, D. H. (Ed.). (1996). *Handbook of research for educational communications and technology.* New York: Macmillan.

King, A. (1992). Facilitating elaborative learning through guided student-generated questioning. *Educational Psychologist, 27,* 111–126.

Lester, J. C., Stone, B., & Stelling, G. (1999). Lifelike pedagogical agents for mixed-initiative problem solving in constructivist learning environments. *User Modeling and User-Adapted Interaction, 9,* 1–44.

Macromedia. (1995a). *Director 4.0.4* [Computer program]. San Francisco: Author.

Macromedia. (1995b). *SoundEdit 16* [Computer program]. San Francisco: Author.

Martin, V. L., & Pressley, M. (1991). Elaborative-interrogation effects depend on the nature of the question. *Journal of Educational Psychology, 83,* 113–119.

Mayer, R. E. (2001). *Multimedia learning.* New York: Cambridge University Press.

Mayer, R. E. (2002). *The promise of educational psychology: Vol. 2. Teaching for meaningful learning.* Upper Saddle River, NJ: Prentice Hall.

Mayer, R. E. (2004). Should there be a three strikes rule against pure discovery learning? The case for guided methods of instruction. *American Psychologist, 59,* 14–19.

Mayer, R. E., Mautone, P. D., & Prothero, W. (2002). Pictorial aids for learning by doing in a multimedia geology simulation game. *Journal of Educational Psychology, 94,* 171–185.

Mayer, R. E., & Moreno, R. (2002). Aids to computer-based multimedia learning. *Learning and Instruction, 12,* 107–119.

McKeough, A., Lupart, J., & Marini, A. (Eds.). (1995). *Teaching for transfer: Fostering generalization in learning.* Mahwah, NJ: Erlbaum.

Moreno, R. (2004). Decreasing cognitive load for novice students: Effects of explanatory versus corrective feedback on discovery-based multimedia. *Instructional Science, 32,* 99–113.

Moreno, R., & Mayer, R. E. (2002a). Learning science in virtual reality multimedia environments: Role of methods and media. *Journal of Educational Psychology, 94,* 598–610.

Moreno, R., & Mayer, R. E. (2002b). Verbal redundancy in multimedia learning: When reading helps listening. *Journal of Educational Psychology, 94,* 156–163.

Moreno, R., Mayer, R. E., Spires, H. A., & Lester, J. C. (2001). The case for social agency in computer-based teaching: Do students learn more deeply when they interact with animated pedagogical agents? *Cognition and Instruction, 19,* 177–214.

Prensky, M. (2001). *Digital game-based learning.* New York: McGraw-Hill.

Pressley, M., McDaniel, M. A., Turnure, J. E., Wood, E., & Ahmad, M. (1987). Generation and precision of elaboration: Effects on intentional and incidental learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 13,* 291–300.

Pressley, M., Symons, S., McDaniel, M. A., Snyder, B. L., & Turnure, J. E. (1988). Elaborative interrogation facilitates acquisition of confusing facts. *Journal of Educational Psychology, 80,* 268–278.

Pressley, M., Wood, E., Woloshyn, V., Martin, V., King, A., & Menke, D. (1992). Encouraging mindful use of prior knowledge: Attempting to construct explanatory answers facilitates learning. *Educational Psychologist, 27,* 91–109.

Schauble, L. (1990). Belief revision in children: The role of prior knowledge and strategies for generating evidence. *Journal of Experimental Child Psychology, 49,* 31–57.

Seifert, T. L. (1993). Effects of elaborative interrogation with prose passages. *Journal of Educational Psychology, 85,* 642–651.

Singley, M. K., & Anderson, J. R. (1989). *The transfer of cognitive skill.* Cambridge, MA: Harvard University Press.

Tuovinen, J. E., & Sweller, J. (1999). A comparison of cognitive load associated with discovery learning and worked examples. *Journal of Educational Psychology, 91,* 334–341.

Willoughby, T., Waller, T. G., Wood, E., & MacKinnon, G. E. (1993). The effect of prior knowledge on an immediate and delayed associative learning task following elaborative interrogation. *Contemporary Educational Psychology, 18,* 36–46.

Wittrock, M. C. (1966). The learning by discovery hypothesis. In L. S. Shulman & E. R. Keislar (Eds.), *Learning by discovery: A critical appraisal* (pp. 33–75). Chicago: Rand McNally.

Woloshyn, V., Paivio, A., & Pressley, M. (1994). Use of elaborative interrogation to help students acquire information consistent with prior knowledge and information inconsistent with prior knowledge. *Journal of Educational Psychology, 86,* 79–89.

Wood, E., Pressley, M., & Winne, P. H. (1990). Elaborative interrogation effects on children's learning of factual content. *Journal of Educational Psychology, 82,* 81–91.

Received January 15, 2003
Revision received July 13, 2004
Accepted July 15, 2004 ■