Journal of Educational Psychology
1999, Vol. 91, No. 2, 358–368

0022-0663/99/$3.00

Cognitive Principles of Multimedia Learning: The Role of Modality and Contiguity

Roxana Moreno and Richard E. Mayer
University of California, Santa Barbara

Students viewed a computer animation depicting the process of lightning. In Experiment 1, they concurrently viewed on-screen text presented near the animation or far from the animation, or concurrently listened to a narration. In Experiment 2, they concurrently viewed on-screen text or listened to a narration, viewed on-screen text following or preceding the animation, or listened to a narration following or preceding the animation. Learning was measured by retention, transfer, and matching tests. Experiment 1 revealed a spatial-contiguity effect in which students learned better when visual and verbal materials were physically close. Both experiments revealed a modality effect in which students learned better when verbal input was presented auditorily as speech rather than visually as text. The results support 2 cognitive principles of multimedia learning.

Technological advances have made possible the combination and coordination of verbal presentation modes (such as narration and on-screen text) with nonverbal presentation modes (such as graphics, video, animations, and environmental sounds) in just one device (the computer). These advances include multimedia environments, where students can be introduced to causal models of complex systems by the use of computer-generated animations (Park & Hopkins, 1993). However, despite its power to facilitate learning, multimedia has been developed on the basis of its technological capacity, and rarely is it used according to research-based principles (Kozma, 1991; Mayer, in press; Moore, Burton, & Myers, 1996). Instructional design of multimedia is still mostly based on the intuitive beliefs of designers rather than on empirical evidence (Park & Hannafin, 1994). The purpose of the present study is to contribute to multimedia learning theory by clarifying and testing two cognitive principles: the *contiguity principle* and the *modality principle.*

Cognitive Principles of Multimedia Learning

Contiguity Principle

Mayer and Anderson (1992) have proposed the instructional design principle called the contiguity principle, which states that the effectiveness of multimedia instruction increases when words and pictures are presented contiguously in time or space. In this article, we distinguish between two effects that can be expected from the contiguity principle: a *temporal-contiguity effect* and a *spatial-contiguity effect.*

A spatial-contiguity effect refers to learning enhancement when printed text and pictures are physically integrated or close to each other rather than physically separated. For example, students who read a booklet explaining how tire pumps work that included captioned illustrations placed near the text generated about 75% more useful solutions on problem-solving transfer questions than did students who read the same text and illustrations presented on separate pages (Mayer, 1989; Mayer, Steinhoff, Bower, & Mars, 1995). In a review of 10 studies concerning whether multimedia instruction is effective, Mayer (1997) concluded that there was consistent evidence for a contiguity effect. Students generated a median of over 50% more creative solutions to transfer problems when verbal and visual explanations were integrated than when they were separated (Mayer, 1997). Similar patterns have been noted by other researchers (Chandler & Sweller, 1991; Paas & Van Merrienboer, 1994; Sweller & Chandler, 1994; Sweller, Chandler, Tierney, & Cooper, 1990).

A temporal-contiguity effect refers to learning enhancement when visual and spoken materials are temporally synchronized, that is, presented simultaneously rather than successively. For example, Mayer and Anderson (1991, 1992) asked a group of students to view an animation depicting the operation of a bicycle pump that included a narration given before or during the animation. The words-with-pictures group outperformed the words-before-pictures group on tests of problem-solving transfer (Mayer & Anderson, 1991). In another study (Mayer & Anderson, 1992), students studied an animation that depicted the operation of bicycle pumps or automobile brakes along with concurrent narration, successive animation and narration, animation alone, narration alone, or no instruction. The temporal-contiguity effect was replicated and extended to new materials. These studies have been cited as evidence for a "temporal example of split attention" (Mousavi, Low, & Sweller, 1995, p. 320). In a related study, Baggett (1984) presented film and voice-over that were in synchrony or shifted relative to the other by 7, 14, or 21 s. Groups with

Roxana Moreno created the multimedia materials used in Experiments 1 and 2.

Correspondence concerning this article should be addressed to Roxana Moreno or Richard E. Mayer, Department of Psychology, University of California, Santa Barbara, California 93106. Electronic mail may be sent to moreno@psych.ucsb.edu or to mayer@psych.ucsb.edu.

visuals and voice in synchrony and groups with visuals preceding voice by 7 s recalled more material than the rest of the groups, thus providing additional evidence for the temporal-contiguity effect for verbal recall.

The contiguity principle has also been described under the name of *split-attention effect* in the cognitive load literature (Chandler & Sweller, 1992; Sweller et al., 1990; Tarmizi & Sweller, 1988). On the basis of research on worked-out examples, Sweller and his colleagues defined the split-attention effect as the impairment in learning that arises from the need to mentally integrate disparate sources of information before instructional material can be rendered intelligible (Chandler & Sweller, 1992; Cooper & Sweller, 1987; Sweller & Cooper, 1985; Ward & Sweller, 1990). Split-attention effects were also produced in instructional studies about numerical control programming (Sweller et al., 1990), a paper folding task (Bobis, Sweller, & Cooper, 1993), and reading scientific reports (Chandler & Sweller, 1992). Finally, Sweller and Chandler (1994) and Chandler and Sweller (1992) extended the split-attention effect to the case where students learned better from a self-contained computer programming manual (which physically integrated all the material and did not require the use of the computer hardware) than from instructions that involved continual interaction between a manual and a computer.

Modality Principle

According to the modality principle, when giving a multimedia explanation, words should be presented as auditory narration rather than as visual on-screen text; that is, words should be presented auditorily rather than visually. Classic verbal learning research on short-term memory provides evidence for a modality effect; that is, "the finding that, in short-term memory tasks, auditory presentation almost always resulted in higher recall than did visual presentation" (Penney, 1989, p. 398). In an extensive review, Penney (1989) considered research in which the presentation material was a mixture of the auditory and visual modalities and suggested that the effective capacity of working memory can be increased by using both visual and auditory channels. One of the most cited examples of modality effects in the literature is the superior recall for lists of items when they are presented in visual and auditory modalities rather than just in one modality (Frick, 1984). Many studies also show that attention can be better divided between the eye and ear than between two auditory or two visual channels (Wickens, 1984). This result would be expected for the cases in which the two visual sources are spatially separated so both cannot access foveal vision simultaneously or in which the two auditory sources mask each other. Nevertheless, the modality effect persists even when the studies control for such peripheral interference (Isreal, 1980; Rollins & Hendricks, 1980; Treisman & Davies, 1973).

Similar to the case of short-term memory, a modality effect can be observed in multimedia learning if students who study from visual presentations with narration outperform students who study the same visual presentation with text. This is the case of a study conducted by Mousavi, Low, and Sweller (1995) involving the presentation of worked-out geometry examples. In the first two experiments, students were presented with worked-out geometry examples under three different conditions. One group of students viewed a diagram and its associated statements and heard the statements played from a tape recorder (simultaneous group). A second group viewed the diagram with the statements (visual–visual group), and the third group viewed the diagram while listening to the statements from a tape recorder (visual–auditory group). Learning was comparatively enhanced by the auditory mode of presentation in both experiments. In two further experiments, the visual and verbal presentations were made sequential, and still, the mixed modality of presentation was superior. Finally, the modality effect was also observed when both sources of information consisted of statements rather than a diagram and statements. The results suggest that dual-presentation modalities may increase working memory resources by activating both auditory and visual working memory rather than just one (Mousavi, Low, & Sweller, 1995).

In a similar fashion, Mayer and Moreno (1998) found that the superiority of simultaneous narrations and animations over simultaneous text and animations is consistent with a dual-processing model of working memory with separate channels for visual and auditory processing. Students were presented with a computer-generated animation depicting the process of lightning formation (Experiment 1) or the operation of a car's braking system (Experiment 2). For each experiment, one group of students was presented with concurrent descriptive narration, and the other group was presented with concurrent on-screen text. Across both studies, the first group outperformed the second group in a retention test, a matching test, and a transfer test. This result was the first demonstration of a modality effect within the context of multimedia learning with animations, where students learn more effectively when an animation was accompanied by a verbal explanation presented in auditory mode as speech rather than in visual mode as on-screen text.

Distinguishing Modality and Contiguity Principles

In Mayer and Moreno's (1998) study, students who learned with concurrent narration and animations outperformed those who learned with concurrent on-screen text and animations. These results might be interpreted as being due to two different effects: a spatial-contiguity effect and a modality effect. In fact, the superiority of concurrent animation and narration over concurrent animation with on-screen text might be caused by students missing part of the visual information while they are reading the on-screen text (or vice versa), by the more efficient processing from two separate auditory and visual streams (Penney, 1989), or by a combination of both causes.

The goal of Experiment 1 was to distinguish between spatial-contiguity and modality effects in multimedia learning with animations. This is the first study to examine

directly the relative contributions of spatial contiguity and modality to multimedia learning. To do so, the physical proximity of the on-screen text and the animation was manipulated. One group of students had on-screen text that was integrated or physically close to the animation (IT group), and a second group of students had on-screen text that was separated or physically far from the animation (ST group). A third group of students saw a presentation with concurrent animation and narration (N group). In this manner, any performance differences between the text groups (IT and ST) can be interpreted exclusively in terms of spatial-contiguity, and any differences in the performance of the narration group (N) relative to the text groups (IT and ST) can be interpreted exclusively in terms of modality.

Experiment 2 was designed to distinguish between temporal-contiguity and modality effects on learning. This is the first study to examine directly the relative contributions of temporal contiguity and modality to multimedia learning. Concurrent multimedia presentations, such as the ones used in Experiment 1, force the text groups to hold material from one source of information (verbal or nonverbal) in working memory before attending to the other source. Therefore, the narration group might have the advantage of being able to attend to both simultaneously, causing a modality effect. This difference might disappear if researchers use sequential multimedia presentations, where verbal and nonverbal materials are presented one after the other. Sequential multimedia presentations force the narration groups to hold material from one source of information (verbal or nonverbal) in working memory before attending to the other. If the advantage of narration over on-screen text resides in a modality principle, then the advantage for auditory–visual presentations should not disappear when they are made sequential. Experiment 2 examined six conditions. First, and similar to Experiment 1, one group of students viewed concurrently on-screen text while viewing the animation (TT), and a second group of students listened concurrently to a narration while viewing the animation (NN). In addition to the concurrent groups, four groups of sequential presentations were included. Students listened to a narration preceding the corresponding portion of the animation (NA), listened to the narration following the animation (AN), read the on-screen text preceding the animation (TA), or read the on-screen text following the animation (AT).

By combining concurrent and sequential presentations with auditory and visual modalities, this design allows comparing and contrasting the two effects. First, a modality effect can be measured by comparing the groups that receive text and animations (TT, AT, TA) with the ones that receive narration and animations (NN, AN, NA). Second, a temporal-contiguity effect can be measured within each modality by comparing the groups that concurrently receive verbal and visual presentations with the ones that sequentially receive verbal and visual presentations (NN vs. NA and AN; TT vs. TA and AT). Experiment 2 extends previous research (Mousavi, Low, & Sweller, 1995) on mixed-modality presentations with diagrams and statements to the context of multimedia learning with animations.

Experiment 1

The purpose of Experiment 1 was to contribute to a cognitive theory of multimedia learning by examining how the spatial contiguity of text and animations and modality affect learning. By manipulating the proximity of the on-screen text with the animation and the modality of verbal information, it is possible to distinguish between split-attention and modality interpretations for differences in performance between groups. Although a split-attention interpretation predicts that students in the IT group should outperform students in the ST group because of the higher cognitive load of the latter with respect to the former, a modality interpretation predicts that students in the N group should outperform both on-screen text groups because of the increase of effective working memory created by mixed modality presentations.

Method

Participants and design. The participants were 132 college students recruited from the psychology subject pool at the University of California, Santa Barbara. All participants were classified as having low experience in meteorology on the basis of a questionnaire. Forty-one participants served in the N group, 41 in the IT group, and 40 participants served in the ST group. Because previous studies have demonstrated that split-attention and modality effects were stronger for low-experience learners than for high-experience learners, we included only low-experience students in our study. We computed an experience score by tallying the number of domain-related items that the participant checked on the knowledge checklist and adding that number to the level of knowledge the participant checked on the 5-level self-assessment (with *very little* counted as 1 point through *very much* as 5). We eliminated the data for any student who scored above 6 and replaced it with the data of a new student. Using this procedure, 10 students were replaced in Experiment 1.

Materials and apparatus. For each participant, the paper-and-pencil materials consisted of a questionnaire, a retention test, a matching test, and a 4-page transfer test, with each typed on 8.5-×-11-inch sheets of paper.[1] The questionnaire solicited informa-

[1] After these tests, participants took a battery of spatial ability tests consisting of an 80-item mental rotation task and a 10-item paper folding test as used in previous research (Mayer & Sims, 1994). For each test, participants read instructions that included a worked-out example and then had 3 min to complete as many test items as possible. A composite spatial ability score was computed for each participant by finding the number correct on the mental rotation test divided by 8 and rounded to the nearest integer and adding that number to the number correct on the mental rotation test (yielding a maximum score of 20). For purposes of data analyses, students were classified as low or high spatial ability on the basis of a median split. Separate two-way ANOVAs with treatment group and spatial ability as between-subjects factors failed to reveal significant interactions on the retention test, $F(2, 116) = 0.12$, $MSE = 1.05$, $p = ns$, matching test, $F(2, 116) = 1.81$, $MSE = 4.03$, $p = ns$, or transfer test, $F(2, 116) = 0.06$, $MSE = 0.11$, $p = ns$. Thus, there is no evidence that spatial ability affected the size of the spatial-contiguity or modality effects for Experiment 1. Individual differences in spatial ability, SAT score, and gender were not the main focus of the study, and these factors were not part of the research design.

tion concerning the participant's SAT scores, gender, and meteorology knowledge. Meteorology knowledge was assessed by using a 7-item knowledge checklist and a 5-item self-rating. The checklist consisted of instructions to "please place a check mark next to the items that apply to you" followed by a list of seven items: "I regularly read the weather maps in the newspaper." "I know what a cold front is." "I can distinguish between cumulous and nimbus clouds." "I know what a low pressure system is." "I can explain what makes the wind blow." "I know what this symbol means: [symbol for cold front]." "I know what this symbol means: [symbol for warm front]." On the 5-item self-rating, students were asked to "please put a check mark indicating your knowledge of meterology (weather):" on a 5-point scale ranging from *very little* (1) to *very much* (5).

The retention test contained the following instructions at the top of the sheet: "Please write down an explanation of how lightning works." The transfer test consisted of the following four questions, each typed on a separate sheet: "What could you do to decrease the intensity of lightning?" "Suppose you see clouds in the sky, but no lightning. Why not?" "What does air temperature have to do with lightning?" "What causes lightning?" The matching test presented four frames from the animation along with the following instructions: "Circle cool moist air and write C next to it. Circle the warmer surface and write W next to it. Circle the updraft and write U next to it. Circle the freezing level and write F next to it. Circle the downdraft and write D next to it. Circle the gusts of cool wind and write G next to it. Circle the stepped leader and write S next to it. Circle the return stroke and write R next to it."

The computerized materials consisted of three computer programs for multimedia presentations on how the lightning process works. The intended audience for the application is a high school or college student who individually is learning a cause-and-effect system, such as looking up an entry in a multimedia encyclopedia. All program versions generated an identical animation depicting air moving from the ocean to the land, water vapor condensing to form a cloud, the rising of the cloud beyond the freezing level, the formation of crystals in the cloud, the movement of updrafts and downdrafts, the building of electrical charges within the cloud, the division of positive and negative charges, the traveling of a negative stepped leader from the cloud to the ground, the traveling of a positive stepped leader from the ground to the cloud, the negative charges following the path to the ground, the meeting of the negative leader with the positive leader, and the positive charges following the path toward the cloud. The N version also included concurrent narration describing each of the major events in words spoken at a slow rate by a male voice. The ST and IT versions included concurrent text presented on the screen using the same words and timing as the narration used in the N group; for the ST version, the text was physically far from the animation, and for the IT version, the text was physically close to the relevant part of the animation. The three versions had an identical total duration of 180 s with the on-screen text and narration displayed for the same amount of time. As can be seen from the first frame of Figure 1, the IT group concurrently viewed integrated text with the animation; that is, the text was physically close to where the animation was taking place. The second frame of Figure 1 shows the same animation frame for the ST group. In this case, students concurrently viewed separated on-screen text with the animation; that is, the text was placed at the bottom of the screen, physically far from where the animation was taking place. The third frame of Figure 1 shows the same animation frame for the N group, who viewed the animation while listening to concurrent narration with no on-screen text. The multimedia presentations were developed using Director *5 and Soundedit 16 (Macromedia, 1997).*

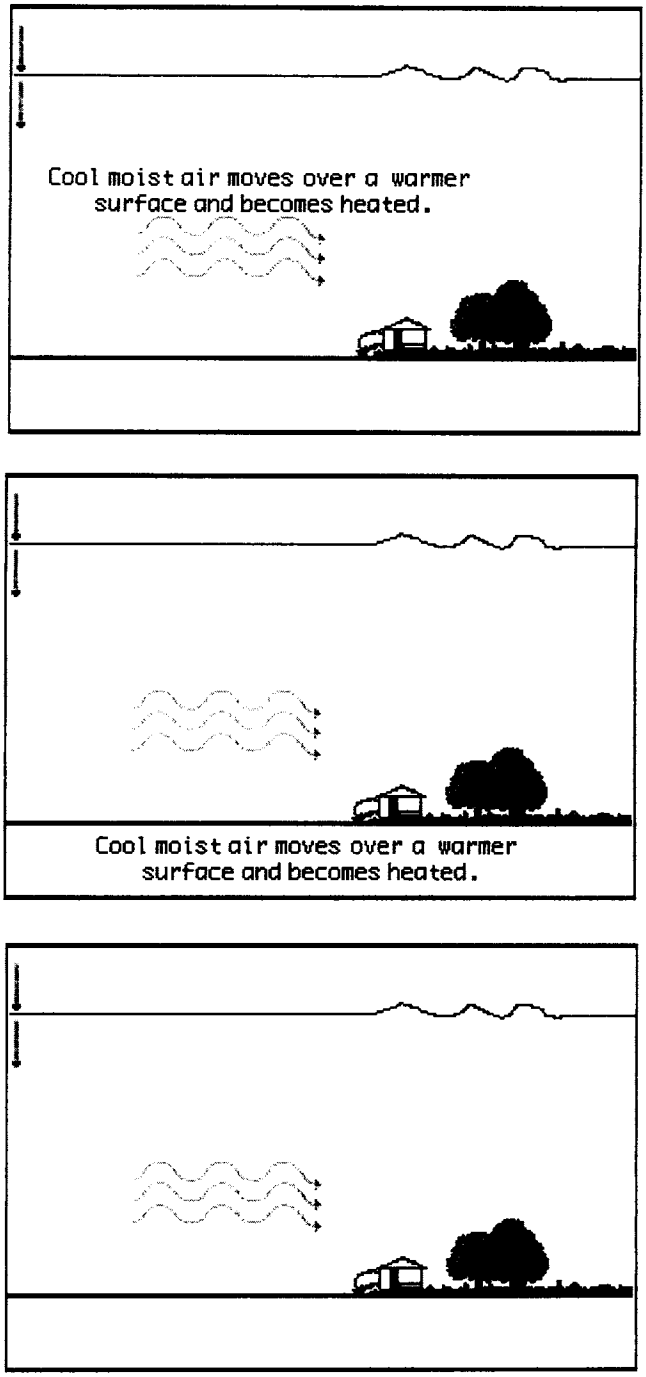

Figure 1. A frame selected from the animation corresponding to the IT, ST, and N groups (top, middle, and bottom panels, respectively). IT = group that had on-screen text integrated or physically close to the animation; ST = group that had on-screen text separated or physically far from the animation; and N = group that saw a presentation with concurrent animation and narration.

The apparatus consisted of 5 Macintosh IIci computer systems, which included a 14-inch monitor and Sony headphones.

Procedure. Participants were tested in groups of 1 to 5 per session. Each participant was randomly assigned to a treatment

group (IT, ST, or N) and was seated at an individual cubicle in front of a computer.

First, participants completed the questionnaire at their own rates. Second, the experimenter presented oral instructions stating that the computer would show an animation of how the process of lightning works and that when the computer was finished the experimenter would have some questions for the participants to answer. Participants in the N treatment were told to put on headphones, and all participants were told to press the space bar to begin the presentation. Third, upon pressing the space bar, the respective version of the animation was presented once to all participants. Fourth, when the presentation 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 along with instructions to write down an explanation of how lightning works. After 5 min, the recall sheet was collected. Then, the 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 matching test was presented and collected after 3 min.

Scoring. A scorer who was not aware of the treatment condition of each participant determined the retention score, matching score, and transfer score for each participant. A second rater scored a randomly picked subset of 20% of the tests. Agreement between both scorers was 98% on the retention tests, 94% on the transfer tests, and 100% on the matching tests. Disagreements were resolved by consensus.

A retention score was computed for each participant by counting the number of major idea units (out of 19 possible) that the participant produced on the retention test. One point was given for correctly stating each of the following 19 idea units regardless of wording: (a) cool air moves, (b) it becomes heated, (c) it rises, (d) water condenses, (e) the cloud extends beyond the freezing level, (f) crystals form, (g) water and crystals fall, (h) it produces updrafts and downdrafts, (i) people feel the gusts of cool wind before the rain, (j) electrical charges build, (k) negative charges fall to the bottom of the cloud (or positive charges go to the top), (l) a step leader travels down, (m) in a step fashion, (n) the leaders meet, (o) at 165 feet from the ground, (p) negative charges rush down, (q) they produce a light that is not very bright, (r) positive charges rush up, and (s) this produces the bright light people see as a flash of lightning.

A matching score was computed for each participant by counting the number of correctly labeled elements (out of eight possible) on the matching test. Participants received 1 point for each part that was circled and labeled with the appropriate letter.

We computed a transfer score for each participant by counting the number of acceptable answers that the participant produced across the four transfer problems. For example, acceptable answers for the first question about how to decrease lightning intensity included removing positive ions from the ground or reducing the temperature difference between the ocean and the earth; acceptable answers for the second question about why could there be clouds but no lightning included stating that the tops of the clouds might not be high enough to freeze or that positive and negative charges might not have built up yet; acceptable answers for the third question about how is temperature related to lightning included stating that the air must be cooler than the ground or that the temperature has to be low enough for the cloud's top to freeze; acceptable answers for the fourth question about what causes lightning included the difference in electrical charges within the cloud or the difference in temperature between the top and the bottom of the cloud. Unacceptable answers for the first question included removing trees and tall objects from the ground; unacceptable answers for the second question included stating that the cloud was not a rain cloud; unacceptable answers for the third question included stating that warm air rises; unacceptable answers for the fourth question included describing the animation step by step without specifying that the difference in charges or temperature were the actual cause. Questions were open ended, so participants could receive as many points per problem as correct answers they gave.

Results

An analysis of variance (ANOVA) was conducted for each dependent measure with group (N, IT, and ST) as the between-subjects factor and with recall, transfer, and matching scores as the respective dependent measure. Newman-Keuls tests were conducted for all cases where the ANOVA yielded a significant effect ($\alpha = .05$).

Effects on verbal recall. The left panel of Figure 2 shows the proportion of the 19 idea units correctly recalled by students in each of the three groups. As can be seen, the N group averaged about 61%, whereas the IT and the ST groups averaged about 48% and 41%, respectively. The ANOVA revealed a main effect for treatment group, $F(2, 119) = 16.137, MSE = 8.861, p < .001$. The N group scored significantly higher than the IT and ST groups, yielding an effect size of 1.00 for narration, and the IT group scored significantly higher than the ST group, yielding an effect size of 0.47 for spatial contiguity. These results are consistent with both the modality interpretation and the spatial-contiguity interpretation.

Effects on problem-solving transfer. As can be seen from the middle panel of Figure 2, the proportion of correct solutions on the transfer test by students in the N, IT, and ST groups was on average 38%, 23%, and 16%, respectively. This proportion was based on a total possible score of 9,

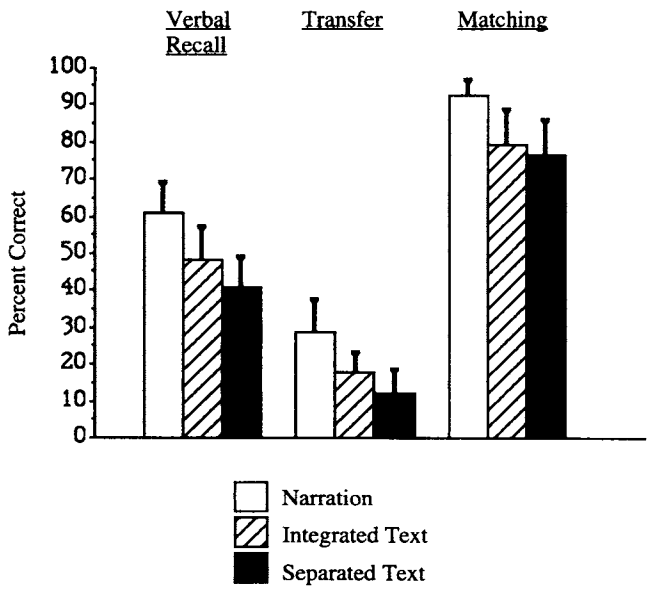

Figure 2. Proportion correct on recall, transfer and matching tests in Experiment 1.

which was the maximum transfer score obtained by any participant in the study. The ANOVA revealed a main effect for treatment group, $F(2, 119) = 24.073$, $MSE = 1.746$, $p < .001$. The N group scored significantly higher than the IT and ST groups, and the IT group scored significantly higher than the ST group. The effect size was 1.06 for modality (N vs. IT and ST) and 0.48 for spatial contiguity (IT vs. ST). These results support both the modality interpretation and the spatial-contiguity interpretation.[2]

Effects on visual–verbal matching. The right panel of Figure 2 shows the proportion correct on the 8-item matching test by students in each group. As can be seen, the N group averaged about 93%, whereas the IT group and the ST groups averaged about 80% and 77%, respectively. The ANOVA revealed a main effect for treatment group, $F(2, 119) = 18.632$, $MSE = 8.203$, $p < .001$. The N group scored significantly higher than the IT and ST groups, which did not differ significantly from each other. The effect size was 1.32 for modality (N vs. IT and ST) and 0.17 for spatial contiguity (IT vs. ST). Students' matching scores were classified as low or high on the basis of a median split. Chi-square tests (with Yates correction) revealed that the proportion of low-scoring students in the text groups (IT and ST) was significantly greater than the proportion of low-scoring students in the narration group, $\chi^2(1, N = 122) = 141.38$, $p < .05$, but the proportion of low-scoring students in the ST group did not differ significantly from the proportion of low-scoring students in the IT group, $\chi^2(1, N = 81) = .11$, $p = ns$. These results support the modality interpretation but not the spatial-contiguity interpretation.

Overall, modality effects were obtained on the retention, matching, and transfer tests, whereas spatial-contiguity effects were obtained on the retention and transfer tests. According to the modality interpretation, students who learn with auditory verbal materials plus animations recall more, solve problems better, and are better able to match the visual and verbal elements than those who learn with on-screen text plus animations. According to the spatial-contiguity interpretation, students presented with integrated on-screen text plus animations recall more and solve problems better than those presented with on-screen text separated from animations. One possible explanation for not finding a significant difference in the matching scores of the IT and ST groups is a ceiling effect in which all groups did very well on this particular measure. The results of Experiment 1 provide support for both modality and spatial contiguity as important but separable factors in multimedia learning.

Experiment 2

Experiment 1 demonstrated the beneficial effects of constructing multimedia presentations in which verbal material is presented auditorily (as speech) rather than visually (as on-screen text). This finding can be interpreted two ways. First, it may suggest that effective working memory capacity is increased if both auditory and visual working memory is used. However, because the on-screen text and the animation were presented simultaneously, students are likely to read a statement, hold it in working memory, and then attend to the animation to integrate it with the text. Despite the fact that attention shifts are very fast, it is not possible to attend to both materials at the same time, and part of one type of material will be lost while the student is attending to the other. This is not the case for the narration group, which can be attending to both materials simultaneously without competition. Therefore, introducing sequential presentations allows us to distinguish between modality and contiguity effects. Forcing the text and narration groups to hold the verbal material either before or after viewing the animation eliminates any differences in the processing that might be due exclusively to the visual attention competition between simultaneous on-screen text and visual information, that is, spatial-contiguity effects. If there is a modality effect, the advantage of auditory verbal input should remain when the presentations are made sequential. Experiment 2 was designed to examine the modality effect by presenting a computer-generated animation depicting the process of lightning formation with either simultaneous or sequential presentation conditions. The simultaneous presentations were identical to the narration group and separated text group from Experiment 1 (now called NN and TT, respectively). The sequential presentations consisted of the AN, NA, AT, or TA version.

Method

Participants and design. The participants were 127 college students recruited from the psychology subject pool at the University of California, Santa Barbara. All participants were classified as low experience in meteorology on the basis of a questionnaire. Twenty low-experience participants served in the NN, NA, TT, TA, and AT groups, and 18 low-experience participants served in the AN group. Using the same procedure as in Experiment 1, we replaced 9 students who indicated that they possessed high knowledge in meteorology.

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 questionnaire, a retention test, a matching test, and a 4-page transfer test, with each typed on 8.5-×-11-inch sheets of paper.

The computerized materials consisted of six computer programs for multimedia presentations on how the lightning process works, using the same animation and words as in Experiment 1. The NN version was identical to the N version for Experiment 1, and the TT version was identical to the ST version for Experiment 1. Both concurrent versions had an identical total duration of 180 s with the on-screen text and narration displayed by the same amount of time. The sequential versions of Experiment 1 consisted of the same animation but divided into 16 cycles, as shown in the appendix. The

[2] We also conducted an analysis of covariance on the problem-solving transfer scores by using recall scores as a covariate. The main effect of treatment group remained significant for Experiment 1, $F(2, 116) = 3.840$, $p < .05$, but not for Experiment 2, $F(1, 114) = 0.656$, $p = ns$. However, there is a problem in interpreting this analysis as the dependent measures were not designed for this kind of fine-leveled analysis; that is, the recall test was not designed to evaluate the exact material that is needed for solving various problem-solving items.

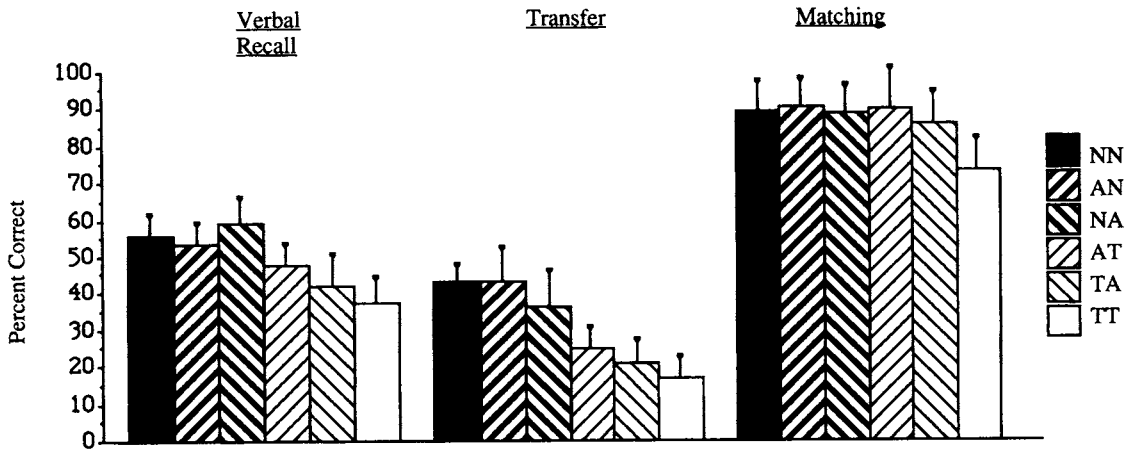

Figure 3. Proportion correct on recall, transfer and matching tests in Experiment 2. Note that NN means the group that received concurrent narration and animation; AN and NA mean the groups that received narration following or preceding the animation, respectively; TT means the group that received concurrent text and animation; and AT and TA mean the groups that received narration following or preceding the animation, respectively.

sequential text versions presented the text in the same place on the computer screen as that of the TT version but either before or after the respective portion of the animation. The AT version consisted of successive cycles of the animation followed by on-screen text. The TA version consisted of successive presentation of on-screen text followed by animation. The sequential narration versions played the same narration presented for group NN either before or after the respective portion of the animation. The AN version consisted of successive cycles of the animation followed by narration, and the NA version consisted of successive presentations of narration followed by animation. The four sequential versions had identical total durations of 300 s each, with the on-screen text and narration being displayed for the same amount of time. We developed the multimedia programs by using Director 5 and Soundedit 16 (Macromedia, 1997).

Procedure. The procedure was identical to that used in Experiment 1 except that each participant was assigned to either group (NN, AN, NA, TT, TA, or AT), and participants in the NN, NA and AN treatments were told to put on headphones before pressing the space bar to begin the presentation.

Scoring. The retention score, matching score, and transfer score for each participant were computed identically to the way they were computed in Experiment 1.

Results

First, a separate ANOVA was conducted for each dependent measure with group (NN, AN, NA, TT, TA, and AT) as the between-subjects factor and with recall, transfer, and matching scores as the respective dependent measure. Newman-Keuls tests were followed for all cases where the ANOVA yielded a significant effect ($\alpha = .05$). Second, a two-factor ANOVA was also conducted for each dependent measure with modality (narration vs. text) and temporal contiguity (simultaneous vs. sequential) as the between-subjects factor and with recall, transfer, and matching scores as the respective dependent measure. The two sequential narration groups (NA and AN) and the two sequential text groups (TA and AT) were collapsed for such purposes.[3]

Effects on verbal recall. The left panel of Figure 3 shows the proportion of idea units correctly recalled by students in each of the six groups. The TT group averaged the lowest with about 38%, followed by the TA, AT, AN, NN, and NA groups, which averaged approximately 42%, 48%, 54%, 56%, and 59%, respectively. The first ANOVA revealed a treatment effect, $F(5, 112) = 5.49$, $MSE = 9.3$, $p < .001$, with the narration groups not differing from each other but the TT and TA groups scoring significantly less than each of the three narration groups. Group AT also scored significantly less than the NN and AN groups but did not differ significantly from group NA. These results reflect a modality effect for verbal recall.

The two-factor ANOVA revealed a main effect for modality, $F(1, 114) = 15.58$, $MSE = 217.3$, $p < .001$, with a mean

[3] When visuals and words have to be presented sequentially, such as in the case of diagrams with narrations, animations with narrations, or video with audio, an order effect would arise if learning is enhanced by a particular order of presentation of the verbal and nonverbal materials. No evidence was found for an order effect in Experiment 2. For each of the dependent measures, post hoc tests failed to show any differences between the sequential groups for each modality. Within the text groups, the AT and TA groups did not differ in the retention tests (Ms = 9.1 and 7.9, SDs = 2.61 and 3.93, respectively), matching tests (Ms = 7.1 and 6.6, SDs = 1.41 and 1.23, respectively), or transfer tests (Ms = 2.2 and 1.9, SDs = 1.36 and 1.17, respectively). Within the narration groups, the AN and NA did not differ in the retention tests (Ms = 10.17 and 11.15, SDs = 2.38 and 3.45, respectively), matching tests (Ms = 7.22 and 6.95, SDs = 0.88 and 0.95, respectively), or transfer tests (Ms = 3.72 and 3.25, SDs = 1.27 and 2.07, respectively). This is the reason why we decided to collapse groups AT and TA as a sequential text group and groups AN and NA as a sequential narration group.

number of ideas recalled of 10.67 and 8.03, respectively, for the narration and text groups (*SD*s = 2.82 and 3.28, respectively). The effect size for modality was 0.94. There was no main effect for temporal contiguity, $F(1, 114) = 1.46$, $MSE = 13.59$, $p = ns$, indicating that simultaneous and sequential groups did not differ in their performance (*M*s = 8.88 and 9.56, *SD*s = 3.28 and 3.35, respectively). The effect size for temporal contiguity was 0.20. There was no significant interaction between modality and temporal contiguity, $F(1, 114) = 0.23$, $MSE = 2.46$, $p = ns$. Narration and text groups had higher scores for sequential presentations than for simultaneous presentations (*M*s = 10.65, 10.68, 7.10, and 8.50; *SD*s = 2.54, 3.00, 3.00, and 3.35 for the narration-simultaneous, narration-sequential, text-simultaneous, and text-sequential, respectively). These results fail to show evidence of a temporal-contiguity effect.

The failure to find differences between the simultaneous and sequential narration groups seems to contradict the temporal contiguity effect found in some previous studies on animations and narrations. Mayer and Anderson (1991, 1992) found that simultaneous auditory and visual presentations were superior to successive auditory and visual presentations. On the other hand, Mousavi et al., (1995) obtained similar results to those reported in the present experiment. In a study where students studied worked-out geometry examples, the visual and verbal materials were presented sequentially and simultaneously, and performance was compared. In both presentation modes, the mixed modality of visual and verbal information was superior (Mousavi et al., 1995). All these studies can be reconciled if the length of the cycles in the sequential presentations is taken into account. In the Mayer and Anderson (1991, 1992) studies, long blocks of animation and narrations were presented successively such that visual and verbal information could not be held in working memory simultaneously. In the case of Mousavi et al. and the present experiment, the successive presentations of the animation and narration were much shorter, only a line or two at a time, and thus unlikely to overload working memory.

Effects on problem-solving transfer. The middle panel of Figure 3 shows the proportion of correct solutions on the transfer test in each of the six groups (out of a maximum score of 9 as in the first experiment). The TT group averaged the lowest with about 17% correct, followed by the TA, AT, NA, NN, and AN groups, which averaged approximately 21%, 24%, 36%, 41%, and 41%, respectively. The first ANOVA revealed that the groups differed significantly in their problem-solving transfer, $F(5, 112) = 7.54$, $MSE = 2.43$, $p < .001$, with the text groups (TT, AT, and TA) scoring significantly lower than the narration groups (NN, AN, and NA). The text and narration groups did not differ from each other.

The two-factor ANOVA showed a main effect for modality, $F(1, 114) = 36.03$, $MSE = 2.41$, $p < .001$, with students in the narration groups scoring significantly more than students in the text groups (*M*s = 3.55 and 1.87, *SD*s = 1.77 and 1.31, respectively). The effect size was 1.09 for modality. These results reflect a modality effect for transfer. There was no main effect for temporal contiguity, $F(1, 114) = 0.29$, $MSE = 2.41$, $p = ns$, indicating that the simultaneous and sequential groups did not differ in their performance (*M*s = 2.60 and 2.74, *SD*s = 1.96 and 1.66, respectively). The effect size was 0.13 for temporal contiguity. There was not a significant interaction between modality and temporal contiguity either, $F(1, 114) = 1.65$, $MSE = 2.41$, $p = ns$. Students in the simultaneous narration group scored only slightly higher than students in the sequential narration groups (*M*s = 3.70 and 3.47, respectively; *SD*s = 1.87 and 1.74, respectively), whereas students in the text groups scored higher on successive than simultaneous conditions (*M*s = 2.05 and 1.50, respectively; *SD*s = 1.26 and 1.36, respectively).

Effects on visual–verbal matching. The right panel of Figure 3 shows the proportion correct on the matching test in each of the six groups. The TT group averaged the lowest with about 73% correct, followed by the TA, NA, NN, AT, and AN groups, which averaged approximately 82%, 87%, 88%, 89%, and 90%, respectively. The first ANOVA revealed that the groups differed significantly in their visual–verbal matching, $F(5, 112) = 3.25$, $MSE = 1.57$, $p < .01$, with the TT group scoring significantly lower than the rest of the groups, which did not differ from each other. As in Experiment 1, it is possible to interpret the lack of differences among groups as being due to ceiling effects because all groups scored very high on this test.

The two-factor ANOVA showed a main effect for modality, $F(1, 114) = 8.58$, $MSE = 1.58$, $p < .005$, and for temporal contiguity, $F(1, 114) = 4.45$, $MSE = 1.57$, $p < .05$. Students in the narration groups had larger mean scores than students in the text groups (*M*s = 7.07 and 6.52, respectively; *SD*s = 0.87 and 1.59, respectively), and the sequential groups outperformed the simultaneous groups (*M*s = 6.96 and 6.45, *SD*s = 1.55 and 1.15, respectively). The effect size was 0.63 for modality and 0.33 for temporal contiguity. The interaction between modality and temporal contiguity was also significant, $F(1, 114) = 3.96$, $MSE = 1.57$, $p < .05$, and indicated that although the text groups performed significantly better when the presentation was sequential than when it was simultaneous (*M*s = 6.85 and 5.85; *SD*s = 1.33 and 1.87, respectively), the sequential and simultaneous narration groups did not differ from each other (*M*s = 7.08 and 7.05, *SD*s = 0.91 and 0.83, respectively). As in Experiment 1, each participant's matching score was classified as high or low based on a median split. Chi-square tests (with Yates correction) revealed that the proportion of low-scoring students in the text groups was significantly greater than the proportion of low-scoring students in the narration groups $\chi^2(1, N = 118) = 5.76$, $p < .05$, but that the proportion of low-scoring students in the sequential groups did not differ significantly from the proportion of low-scoring students in the simultaneous groups, $\chi^2(1, N = 37) = 0.45$, $p = ns$. Overall, these results can be interpreted as demonstrating the negative effects of presenting two competing visual materials simultaneously, which is called the split-attention effect in the cognitive load literature (Chandler & Sweller, 1992).

Overall, modality effects were obtained on the retention, transfer, and matching tests, yielding consistent evidence to support the presentation of verbal materials in an auditory

modality. No evidence for temporal-contiguity effects was found within the narration groups. Nevertheless, a temporal-contiguity effect was found within the text groups for verbal and visual matching, where a significant impairment occured when text and animations were presented concurrently as opposed to sequentially.

General Discussion

Implications

The present study has important theoretical and practical implications. Theoretically, it contributes to multimedia learning theory by clarifying and testing two cognitive principles: the contiguity principle and the modality principle. First, it examines whether the spatial-contiguity effects obtained with illustrations in text would also be observed in the context of multimedia learning involving animations and on-screen text. Experiment 1 is the first study to directly examine the relative contributions of spatial contiguity and modality to multimedia learning. The contribution of spatial contiguity to multimedia learning is shown by the finding that learning is impaired when on-screen text is spatially separated from the visual materials, consistent with Baddeley's (1992) model of limited resources in visual working memory and previous research on illustrations in text (Mousavi et al., 1995). The contribution of modality to multimedia learning is shown by the finding that mixed-modality presentations are superior to the most integrated text and visual presentations, consistent with Paivio's (1986) theory that when learners can concurrently hold words in auditory working memory and pictures in visual working memory, they are better able to devote attentional resources to building connections between them.

Second, this research provides the first study to examine the relationship between temporal-contiguity and modality to multimedia learning in multimedia learning with animations. In Experiment 2, we tested the hypothesis that the superiority of concurrent narration and animations found in Experiment 1 was due to the fact that concurrent on-screen text and animations require students to hold one source of the visual materials in working memory while attending to the other source. In Experiment 2, the advantage of narration presentations over on-screen text presentations did not disappear when both groups were forced to hold the information presented in one modality before attending to the other. On the basis of these results, we conclude not only that more information is likely to be held in both auditory and visual working memory rather than in just one but also that the combination of auditory verbal materials with visual nonverbal materials may create deeper understanding than the combination of visual verbal and nonverbal materials. In sum, the results of Experiment 2 are consistent with the idea of independent visual and auditory processors in working memory and tend to extend past evidence for modality effects in short-term memory (Penney, 1989) to a higher cognitive process, namely, learning.

On the practical side, the present study has direct implications for display design and display formatting. In the design of multimedia presentations, the designer is faced with the need to compress the materials into a relatively small area of the student's visual field: the computer screen. We found in Experiment 1 that the degree of closeness of the elements within the display affects students' learning. The excessive scanning that results from placing the text far from where the animation takes place impaired performance significantly. Therefore, when designing instructional software, it is crucial to physically integrate the corresponding graphic and text materials in a multimedia lesson as much as possible, rather than separating them out for reasons of aesthetics.

In Experiment 2, we found that replacing the text with narration eliminates the competing verbal and nonverbal visual inputs. Interestingly, the advantage of the narration over text did not disappear when presentations were made sequential. These results suggest that it is better to present the learner with materials that allow for processing corresponding text and graphics in parallel, that is, by using mixed modality presentations.

Clarifying Cognitive Principles of Multimedia Design

The present study points out the need for clarity in the definition and nomenclature of the cognitive principles involved in multimedia learning. The names chosen to identify such principles become essential for theoretical progress in the field. As research progresses in the area of multimedia learning, the need for a more specific nomenclature is starting to become obvious. For example, the term *split-attention effect* is starting to present theoretical problems as to what type of cognitive effects it refers to. In the past, split-attention effect has been used to refer to either spatial-contiguity effects, temporal-contiguity effects, or modality effects (Sweller & Chandler, 1994). In addition to the nonspecificity of the term split attention, its usage presents the additional problem of suggesting that attention has a unitary nature or architecture. This position, although supported originally by traditional theories on attention, is now being abandoned in view of more recent findings in the neuropsychological area (Allport, 1993).

To address these problems, we propose replacing the general term split-attention effect, which refers to several effects of very different cognitive nature, with the terms spatial-contiguity effect, temporal-contiguity effect, or modality effect, according to the appropriate case. These terms present two benefits. First, they are more clear in describing the empirical finding that they are related to. Spatial contiguity clearly indicates effects in learning because of differences in the closeness of the verbal and nonverbal visual materials that need to be integrated in a lesson. Temporal contiguity clearly refers to the effects in learning that arise because of differences in synchronicity between the verbal and nonverbal materials that need to be integrated in a lesson. Modality clearly indicates the effects in learning that arise from using different modalities to represent verbal materials in a multimedia lesson. Most importantly, the

proposed terminology is not indicative of a particular position in the controversial literature on the unitary nature of attention but rather is open to new findings in the field of attention.

The computer applications used in Experiments 1 and 2 are examples of materials that students can use in the classroom as reference tools. The intended audience is high school or college students who are learning a cause-and-effect system individually with computer-generated animation. The purpose of the present study was not to compare individual learning to cooperative learning (such as learning in pairs or trios of students) or to advocate for students' isolated use of such materials (without the guidance of the teacher). Instead, the purpose was to help determine research-based principles for how to design multimedia materials, such as entries in a multimedia encyclopedia, that can complement other educational activities in a classroom context.

Finally, it is important to point out that the conclusions drawn from the present study might vary widely depending on individual differences. Prior studies have shown that spatial ability (Mayer & Sims, 1994), information coordination ability (Yee, Hunt, & Pellegrino, 1991), and experience (Mayer & Gallini, 1990) can modify to some extent the effects found above. We did not focus on individual differences in spatial ability, coordination ability, or experience in this research.

Only low-experience students were studied. On the basis of past research, high-experience students would have been less likely to exhibit the spatial-contiguity effect and modality effect that we found (Mayer, 1997). If presenting verbal information in an auditory mode allows students to increase their effective working memory capacity, low-experience students who lack a mental model for the instructional material would be the ones to benefit the most from having more cognitive resources available. Additional research is needed to determine the role of individual differences in multimedia learning.

References

Allport, A. (1993). Attention and control: Have we been asking the wrong questions? A review of 25 years. *Attention & Performance, 14,* 183–218.

Baddeley, A. (1992). Working memory. *Science, 255,* 556–559.

Baggett, P. (1984). Role of temporal overlap of visual and auditory material in forming dual media associations. *Journal of Educational Psychology, 76,* 408–417.

Bobis, J., Sweller, J., & Cooper, M. (1993). Cognitive load effects in a primary school geometry task. *Learning and Instruction, 3,* 1–21.

Chandler, P., & Sweller, J. (1991). Cognitive load theory and the format of instruction. *Cognition and Instruction, 8,* 293–332.

Chandler, P., & Sweller, J. (1992). The split-attention effect as a factor in the design of instruction. *British Journal of Educational Psychology, 62,* 233–246.

Cooper, G., & Sweller, J. (1987). The effects of schema acquisition and rule automation on mathematical problem-solving transfer. *Journal of Educational Psychology, 79,* 347–362.

Frick, R. W. (1984). Using both an auditory and a visual short-term store to increase digit span. *Memory & Cognition, 12,* 507–514.

Isreal, J. (1980). *Structural interference in dual task performance: Behavioral and electrophysiological data.* Unpublished doctoral dissertation, University of Illinois, Champaign.

Kozma, R. B. (1991). Learning with media. *Review of Educational Research, 61,* 179–211.

Macromedia. (1997). *Director 5.0* [Computer program]. San Francisco: Author.

Mayer, R. E. (1989). Systematic thinking fostered by illustrations in scientific text. *Journal of Educational Psychology, 81,* 240–246.

Mayer, R. E. (1997). Multimedia learning: Are we asking the right questions? *Educational Psychologist, 32,* 1–19.

Mayer, R. E. (in press). Instructional technology. In F. Durso (Ed.), *Handbook of applied cognition.* Chichester, England: Wiley.

Mayer, R. E., & Anderson, R. B. (1991). Animations need narrations: An experimental test of a dual-dual coding hypothesis. *Journal of Educational Psychology, 83,* 484–490.

Mayer, R. E., & Anderson, R. B. (1992). The instructive animation: Helping students build connections between words and pictures in multimedia learning. *Journal of Educational Psychology, 84,* 444–452.

Mayer, R. E., & Gallini, J. K. (1990). When is an illustration worth ten thousand words? *Journal of Educational Psychology, 82,* 715–726.

Mayer, R. E., & Moreno, R. (1998). A split-attention effect in multimedia learning: Evidence for dual processing systems in working memory. *Journal of Educational Psychology, 90,* 312–320.

Mayer, R. E., & Sims, V. K. (1994). For whom is a picture worth a thousand words? Extensions of a dual-coding theory of multimedia learning. *Journal of Educational Psychology, 86,* 389–401.

Mayer, R. E., Steinhoff, K., Bower, G., & Mars, R. (1995). A generative theory of textbook design: Using annotated illustrations to foster meaningful learning of science text. *Educational Technology Research & Development, 43,* 31–43.

Moore, D. M., Burton, J. K., & Myers, R. J. (1996). Multiple-channel communication: The theoretical and research foundations of multimedia. In D. H. Jonassen (Ed.), *Handbook of research for educational communication and technology* (pp. 851–875). New York: Macmillan.

Mousavi, S., Low, R., & Sweller, J. (1995). Reducing cognitive load by mixing auditory and visual presentation modes. *Journal of Educational Psychology, 87,* 319–334.

Paas, F. G. W. C., & Van Merrienboer, J. J. G. (1994). Variability of worked examples and transfer of geometrical problem solving skills: A cognitive load approach. *Journal of Educational Psychology, 86,* 122–133.

Paivio, A. (1986). *Mental representations: A dual coding approach.* Oxford, England: Oxford University Press.

Park, I., & Hannafin, M. (1994). Empirically-based guidelines for the design of interactive multimedia. *Educational Technology Research and Development, 41,* 63–85.

Park, O., & Hopkins, R. (1993). Instructional conditions for using dynamic visual displays: A review. *Instructional Science, 21,* 427–449.

Penney, C. G. (1989). Modality effects and the structure of short-term verbal memory. *Memory & Cognition, 17,* 398–422.

Rollins, R. A., & Hendricks, R. (1980). Processing of words presented simultaneously to eye and ear. *Journal of Experimental Psychology: Human Perception and Performance, 6,* 99–109.

Sweller, J., & Chandler, P. (1994). Why some material is difficult to learn. *Cognition and Instruction, 12,* 185–233.

Sweller, J., Chandler, P., Tierney, P., & Cooper, M. (1990). Cognitive load and selective attention as factors in the structuring of technical material. *Journal of Experimental Psychology: General, 119,* 176–192.

Sweller, J., & Cooper, M. (1985). The use of worked examples as a substitute for problem solving in learning algebra. *Cognition and Instruction, 2,* 59–89.

Tarmizi, R., & Sweller, J. (1988). Guidance during mathematical problem solving. *Journal of Educational Psychology, 80,* 424–436.

Treisman, A., & Davies, A. (1973). Divided attention to eye and ear. In S. Kornblum (Ed.), *Attention and Performance IV.* New York: Academic Press.

Ward, M., & Sweller, J. (1990). Structuring effective worked out examples. *Cognition and Instruction, 7,* 1–39.

Wickens, C. D. (1984). *Engineering psychology and human performance.* Columbus, Ohio: Merrill.

Yee, P. L., Hunt, E. B., & Pellegrino, J. W. (1991). Coordinating cognitive information: Task effects and individual differences in integrating information from several sources. *Cognitive Psychology, 23,* 615–680.

Appendix

Text/Narration Cycles for the Sequential Presentations

1. Cool moist air moves over a warmer surface and becomes heated.
2. Warmed moist air near the earth's surface rises rapidly.
3. As the air in this updraft cools, water vapor condenses into water droplets and forms a cloud.
4. The cloud's top extends above the freezing level, so the upper portion of the cloud is composed of tiny ice crystals.
5. Within the cloud, the rising and falling air currents cause electrical charges to build.
6. The charge results from the collision of the cloud's rising water droplets against heavier, falling pieces of ice.
7. The negatively charged particles fall to the bottom of the cloud, and most of the positively charged particles rise to the top.
8. Eventually, the water droplets and ice crystals become too large to be suspended by updrafts.
9. As raindrops and ice crystals fall through the cloud, they drag some of the air in the cloud downward, producing downdrafts.
10. When downdrafts strike the ground, they spread out in all directions, producing the gusts of cool wind people feel just before the start of the rain.
11. A stepped leader of negative charges moves downward in a series of steps. It nears the ground.
12. A positively charged leader travels up from such objects as trees and buildings.
13. The two leaders generally meet about 165 feet above the ground.
14. Negatively charged particles then rush from the cloud to the ground along the path created by the leaders. It is not very bright.
15. As the leader stroke nears the ground, it induces an opposite charge, so positively charged particles from the ground rush upward along the same path.
16. This upward motion of the current is the return stroke. It produces the bright light that people notice as a flash of lightning.

Received July 23, 1998
Revision Received November 19, 1998
Accepted November 19, 1998 ■