For Whom Is a Picture Worth a Thousand Words?
Extensions of a Dual-Coding Theory of Multimedia Learning

Richard E. Mayer and Valerie K. Sims

In 2 experiments, high- and low-spatial ability ability students viewed a computer-generated animation and listened simultaneously (concurrent group) or successively (successive group) to a narration that explained the workings of either a bicycle tire pump (Experiment 1) or of the human respiratory system (Experiment 2). The concurrent group generated more creative solutions to subsequent transfer problems than did the successive group; this contiguity effect was strong for high- but not for low-spatial ability students. Consistent with a dual-coding theory, spatial ability allows high-spatial learners to devote more cognitive resources to building referential connections between visual and verbal representations of the presented material, whereas low-spatial ability learners must devote more cognitive resources to building representation connections between visually presented material and its visual representation.

The computer technology for implementing an interactive, multimedia encyclopedia or textbook exists today (Frisse, 1988; Raymond & Tompa, 1988). However, in spite of advances in educational technology, the field of educational psychology lacks a corresponding research-based theory of technology of how to design computer-based instruction using words and pictures (Rieber, 1990a). In attempting to close the gap between educational technology and educational theory, we examine how individual differences affect students' learning from visual and verbal instruction (Mayer, 1989a, 1989b, Mayer & Anderson, 1991, 1992; Mayer & Gallini, 1990). In particular, our goal in the present study is to identify the role of the student's spatial ability in learning from words and pictures about how a system works.

Although less research has been conducted on visual learning than on verbal learning, there are many indications of the power of visual instructional aids (Mandl & Levin, 1989; Schnotz & Kulhavy, in press; Willows & Houghton, 1987a, 1987b; Winn, 1991). In particular, an increasing body of research evidence supports the contention that student learning is affected positively by presenting text and illustrations together (Bernard, 1990; Glenberg & Langston, 1992; Guri-Rozenblit, 1988; Purnell & Solman, 1991; Reed & Beveridge, 1986, 1990; Waddill, McDaniel, & Einstein, 1988). Furthermore, computer-generated animation offers a potentially powerful medium for presenting visually based information to learners (Rieber, 1990b, 1991; White, 1984).

We are motivated by the following question: How can we help students use verbal and visual information to understand scientific explanations? By “help” we mean providing an instructional method, such as presenting visual and verbal explanations in close proximity. In our work, the primary independent variable is whether an animation and narration about how a system works are presented concurrently or successively. By “understand” we mean the ability to transfer the material to new situations, such as being able to generate solutions to problems that are based on the presented explanation of the system. In our studies, the primary dependent measure is the number of acceptable solutions that students generate for divergent problem-solving questions; these questions call for troubleshooting the system, suggesting changes in the system to accomplish specific goals, and hypothesizing why a certain event occurs within the system. By “explanation” we mean a description of a causal system containing parts that interact in a coherent way, such as a description of how a pump works or how the human respiratory system works. Unlike some descriptions that lack a coherent structure, explanations are organized as cause-and-effect chains in which a change in one part causes a change in another part. In our studies, the instructional materials involve cause-and-effect explanations of how various systems work.

In this introduction, we describe a dual-coding theory of multimedia learning. Then, we summarize three research issues—the contiguity effect, the role of experience in the contiguity effect, and the role of ability in the contiguity effect. We close with a summary of our predictions concerning the role of spatial ability in learning from words and pictures.

---

We acknowledge the useful suggestions of Jim Clark and Mark Sadoski.

Correspondence concerning this article should be addressed to either Richard E. Mayer or Valerie K. Sims, Department of Psychology, University of California, Santa Barbara, California 93106. Electronic mail may be sent to either mayer@psych.ucsb.edu or sims@psych.ucsb.edu.

---

Dual-Coding Theory of Multimedia Learning

Multimedia learning occurs when students use information presented in two or more formats—such as a visually presented animation and verbally presented narration—to...
construct knowledge. In a strict sense, our definition applies to the term "multimodal" (which refers to the idea that the learner uses more than one sense modality) rather than "multimedia" (which refers to the idea that the instructor uses more than one presentation medium). Visual and verbal processing refer to two different sense modalities; animation and narration refer to two different presentation media. It also should be noted that, under certain circumstances, verbal material can evoke the construction of visual representations, and visual material can evoke the construction of verbal representations.

In Figure 1 we summarize a dual-coding theory of multimedia learning. Our version of dual-coding theory, adapted and modified from Paivio's theory (Paivio, 1971, 1986; Clark & Paivio, 1991), offers a three-process account of how visually and verbally presented material might be integrated within the learner's working memory during learning. On the top left portion of the figure, a verbal explanation, such as an oral narration, is presented to the learner. Within working memory the learner constructs a mental representation of the system described in the verbal explanation. The cognitive process of going from an external to an internal representation of the verbal material is called building a verbal representational connection (or verbal encoding). On the bottom left portion of the figure, a visual explanation is presented to the learner, such as an animation. Within working memory the learner constructs a mental representation of the visually presented system. The cognitive process of going from an external to an internal representation of visual information is called building a visual representational connection (or visual encoding) and is indicated by the second arrow. The third arrow denotes the construction of referential connections between the two mental representations, that is, the mapping of structural relations between the two representations of the system. For example, building referential connections involves noting that a verbal statement such as "the diaphragm moves down making more room for the lungs" is analogous to an animation showing the diaphragm moving down and thereby creating more space for the lungs.

In understanding an explanation of how pumps work, the learner can build referential connections between visual and verbal representations of essential parts, actions, relations, and principles in the system. For example, the top portion (Parts) of Figure 2 shows that the learner needs to build connections between verbal names and the corresponding pictures for each of seven parts of the pumping system. In the explanation of how a pump works, there are ten actions stated in words, each of which corresponds to a movement in the animation. Five of these mappings are shown in the bottom (Actions) portion of Figure 2. The learner must construct connections between causal relations stated in words (such as, "when the piston moves up, the inlet valve opens") and the corresponding visual presentations (such as an animation of the inlet valve opening as the piston moves up). Finally, the learner may construct connections between underlying causal principles in formal form and in visual form (e.g., the formal concept of air pressure as the density of air and the visual images showing air pressure as the amount of space for air in the lower part of the cylinder).

Performance is shown on the right side of the Figure 1, and it includes the learner's response to tests of retention and transfer. We propose that when the student is asked to solve transfer problems, performance depends on all three connections being formed—visual representational connections, verbal representational connections, and referential connections. Although Paivio's (1986) dual-coding theory did not emphasize problem-solving transfer as a dependent measure, our extended version of dual-coding theory allows us to make predictions concerning problem-solving transfer. According to this version of dual-coding theory, instructional methods that promote the formation of all three connections are more likely to promote problem-solving transfer, whereas instructional methods that fail to promote the formation of one or more of these connections are less likely to promote transfer.

Contiguity Effect

What happens when visual and verbal explanations are presented in a coordinated fashion versus when visual and verbal explanations are presented separately? According to the dual-coding theory presented in Figure 1, the learner uses two distinct information-processing systems—one that represents information verbally and one that represents information visually. For meaningful learning that supports problem-solving transfer, the learner must build an internal verbal representation from the presented verbal information, an internal visual representation from the presented visual information, and referential connections between these verbal and visual representations. From this interpretation of dual-coding theory, we predict that students will be better able to build referential connections when verbal and visual materials are presented contiguously than when they are presented separately. Separate presentation of visual and verbal information encourages the
Verbally Based System | Visually Based System
--- | ---
Parts
- air
- handle
- piston
- cylinder
- inlet valve
- outlet valve
- hose
Actions
- The handle is pulled up
- The piston moves up
- The inlet valve opens
- The outlet valve closes
- Air enters the lower part of the cylinder

Figure 2. A structure mapping of visual and verbal representations of a bicycle tire pump.

Role of Experience in Learning From Animations and Narrations

One important characteristic of learners is their prior experience related to the specific domain of the lesson. We define low-experience learners as those who possess small amounts of domain-specific knowledge and high-experience learners as those who possess large amounts of domain-specific knowledge. For example, for a lesson on how pumps work, high-experience students are those who have taken extensive coursework in mechanics or who regularly repair mechanical devices containing valves and pistons, whereas low-experience students have not had these experiences. For a lesson on the human respiratory system, high-experience learners are those who have taken extensive coursework in human anatomy or who have had medical experiences such as knowing how to administer cardiopulmonary resuscitation (CPR), whereas low-experience students have not had these experiences.

What happens to the contiguity effect when the learner possesses large versus small amounts of relevant domain knowledge? Concurrent presentation of animation and narration enables all three conditions necessary for meaningful learning to be met. Therefore, with concurrent presentation we predict similarly high problem-solving transfer for both low- and high-experience learners. When animation and narration are presented successively, however, the two representations are not in memory at the same time, so the construction of referential connections is hampered. For learners who have no means to overcome this obstacle—such as low-experience learners—we predict that problem-solving transfer will be relatively poor. In contrast, we predict no decrease in problem-solving transfer for high-experience students because they may be able to retrieve relevant knowledge from long-term memory as they listen to the narration, as indicated by the fourth arrow in Figure 1; these high-experience learners are more likely to construct a mental image solely from the words that are presented and to referentially link their verbal representation and visual image.

As predicted, Mayer and Gallini (1990) found across three studies that coordination of words and pictures improved problem-solving transfer for low- but not for high-experience learners. We interpret these results to indicate that domain-specific knowledge compensates for uncoordinated instruction. In particular, when a useful visual model is not presented along with a verbally presented explanation, high-knowledge students are more able than are low-knowledge students to retrieve a source model from long-term memory and to use it to help interpret the incoming verbal explanation. In the present study, we focus solely on low-experience learners, because they are most likely to benefit from coordinated instruction.
Role of Ability in Learning From Animations and Narrations

A second important characteristic of learners is the cognitive abilities that they bring to the learning situation. For example, learning from words and pictures depends on verbal and spatial abilities, and our focus in this study is on comparing the learning outcomes of students who score low or high on tests of spatial ability. Although spatial ability has been defined in several ways (Carroll, 1993; Cronbach & Snow, 1977; Kosslyn, 1980; Lohman, Pellegrino, Alderton, & Regian, 1987; McGee, 1979; Mumaw & Pellegrino, 1984; Richardson, 1980; Smith, 1964; Thurstone & Thurstone, 1941), we focus on one aspect of spatial ability that seems to be most relevant to learning from animations—spatial visualization. Spatial visualization is the ability to mentally rotate or fold objects in two or three dimensions and to imagine the changing configurations of objects that would result from such manipulations. Sternberg (1990) noted that “this ability is involved in visualizing shapes, rotation of objects, and how pieces of a puzzle would fit together” (p. 93). Thurstone and Thurstone (1941) identified spatial visualization as one of seven primary mental abilities.

Tests of spatial visualization emphasize the mental manipulation of objects or the imagining of movement among objects (Carroll, 1993). For example, in Figure 3 there are example items from a paper folding test and a card rotations test (Ekstrom, French, & Harman, 1976). In a paper folding task, the subject must imagine that a sheet of paper has been folded in a certain way, a hole is punched through all thicknesses of the paper at a certain point, and then the sheet is unfolded. The folding and punching are indicated on the left side of the sheet, and the subject must select which of five unfolded sheets is the result. In a card rotations task, the subject must tell whether two shapes are the same or different; if the shapes would match if one were slid or rotated, then the correct answer is “same”; however, if one of the shapes must be flipped or redrawn, then the correct answer is “different.”

What happens to the contiguity effect when learners have either high- or low-spatial ability? In the present study, we seek to understand the conditions under which students build the referential connections required for meaningful learning of scientific systems. In particular, our goal is to determine the role of spatial information processing skill in learning from animations and narrations.

Following earlier work by Cronbach and Snow (1977), we examine two ways in which spatial ability may affect learning about a mechanical or a biological system. First, like prior experience, spatial ability may compensate for poor instruction. That is, with an unsynchronized presentation, a high-spatial ability learner may be able to keep an image active in working memory in the absence of visual stimuli and to construct referential connections between this image and the verbally based representation. In Figure 1, for example, high-spatial ability learners would be able to construct and hold visual representations in working memory as indicated by arrow 2 even when the visual and verbal explanations are presented successively. According to this ability-as-compensator hypothesis, the high-spatial ability learner should be able to construct referential connections between visual and verbal information (indicated by arrow 3) both when the two types are presented contiguously and when they are presented successively. In contrast, low-spatial ability students should only be able to construct referential connections when visual and verbal information

![Paper Folding](image)

![Card Rotations](image)

are presented simultaneously. If spatial ability compensates for unsynchronized instruction, then there should be a large contiguity effect for low-spatial ability students but not for high-spatial ability students.

Alternatively, high spatial ability may enhance good instruction. Under this hypothesis, neither the high- nor the low-spatial ability learner is able to construct referential connections when verbal and visual information are not presented contiguously. However, when the information is presented contiguously, the ability-as-enhancer hypothesis holds that low-spatial ability learners must devote more cognitive resources to the construction of visual representational connections (cf. arrow 2 in Figure 1) than the high-spatial ability learners, and therefore would have fewer cognitive resources available for the construction of referential connections (cf. arrow 3 in Figure 1). Low-spatial ability learners may have to rerun a mental animation several times as they try to construct a visual representation, thus using cognitive resources that then cannot be devoted to other tasks. Even when low-spatial ability learners are given coordinated instruction that allows for verbally based and visually based representations to be in working memory at the same time, they may be unable to allocate the cognitive resources needed to build referential connections between the two representations. In contrast, high-spatial ability learners should be able to benefit from synchronized instruction, because they can devote extra cognitive resources to the construction of referential connections. If high-spatial ability enhances coordinated visual and verbal instruction, then there should be a large contiguity effect for high-spatial ability students but not for low-spatial ability students.

An alternative explanation is that low-spatial ability learners can form visual representations as easily as high-spatial ability learners, but they simply have more difficulty with the referential step. This study was not designed to distinguish this explanation from the limited resource explanation presented earlier. However, other research shows that people who score low in spatial ability have shorter memory spans for visual information than do people who score high in spatial ability (Carroll, 1993), which somewhat supports the limited resource explanation. More research is needed to determine whether the predicted effects can be attributed to differences in building visual representations or in building connections between visual and verbal representations.

The goal of the present study is to provide empirical data concerning the predictions of the ability-as-compensator and ability-as-enhancer hypotheses, as summarized in Table 1. The predictions are intended to represent relative rather than absolute performance. Instead of predicting that there will be no transfer under certain conditions and complete transfer under others, we predict that transfer will be decreased under certain conditions and increased under others.

We tested these predictions in two experiments by comparing coordinated presentation of narration and animation, uncoordinated presentation, and no presentation. The students were all low-experience learners who scored either low or high in spatial ability. Consistent with previous research on low-experience learners, we predicted a contiguity effect in which students who receive concurrent presentation perform better on tests of problem solving than do students who do not receive concurrent presentation (Mayer & Anderson, 1991, 1992). The focus of these experiments is on whether the contiguity effect is strong for low-spatial ability learners but not for high-spatial ability learners (as predicted by the ability-as-compensator hypothesis) or whether the contiguity effect is strong for high-spatial ability learners but not for low-spatial ability learners (as predicted by the ability-as-enhancer hypothesis).

### Experiment 1

In Experiment 1, low-experience students who scored low and high on measures of spatial ability learned how a bicycle tire pump works.

#### Method

**Participants and design.** The participants were 86 college students who lacked extensive prior knowledge about mechanical devices. Ten high-spatial ability and 12 low-spatial ability students served in the concurrent group, 21 high-spatial ability and 22 low-spatial ability students served in the successive group (with approximately half of the students receiving narration followed by animation and the others receiving animation followed by narration), and 7 high-spatial ability and 14 low-spatial ability students were in the control group.

**Materials and apparatus.** The paper-and-pencil materials consisted of a questionnaire, a mental rotation test, a paper folding test, and a problem-solving test, each typed on 8.5 in. × 11 in. (21.59 cm × 27.94 cm) sheets of paper.\(^1\) On the questionnaire students were asked to rate their knowledge of “how to fix house-

---

\(^1\) Both Experiment 1 and Experiment 2 also contained retention tests in which students were given 5 min (in Experiment 1) or 2.5 min (in Experiment 2) to write down all they could remember from the lesson. However, because problem-solving transfer is the major focus of this study, we have not included a discussion of the retention data in this article. The mean retention scores (and standard deviations) for the concurrent, successive, and control groups in Experiment 1 were 7.3 (2.4), 7.8 (1.9), and 2.3 (1.5)
hold appliances and machines" on a 5-point scale ranging from very little (1) to very much (5), and to "put a check mark next to things you have done" for the following list of six activities (each check mark counted as 1 point): (a) "I own a set of tools including screwdrivers, pliers, and wrenches"; (b) "I own at least one power tool"; (c) "I have replaced the heads in a lawn sprinkler system"; (d) "I have replaced the washer in a sink faucet"; (e) "I have installed plumbing pipes or plumbing fixtures." An experience score was determined by tallying the number of points attained. The paper folding test consisted of 10 items, each similar to the example shown in the top portion of Figure 3; each item visually described a sheet of paper that was folded one or more times with a hole punched through the folded paper and presented five alternative answers for how the sheet would look when it was unfolded. The instructions called for selecting the correct alternative for each problem, and one point was awarded for each correct answer. The mental rotation test consisted of 10 rows of nonsense shapes; each row contained a target shape on the left side and eight rotated shapes on the right side, as shown in the bottom portion of Figure 3. The instructions called for marking each rotated shape as "same" if it was a rotated version of the target and "different" if it was not. A score was determined by subtracting the number of incorrect answers from the total number of correct answers and dividing the result by 8, so the maximum score was 10. The mental rotation and paper folding tests were selected from the Kit of Factor Referenced Tests for Cognitive Factors (French, Ekstrom, & Price, 1963). A composite spatial ability score was determined by adding the scores on the two tests; students scoring less than 16 were classified as low-spatial ability (M = 11.9), and students scoring 16 or above were classified as high-spatial ability students (M = 18.7). The problem-solving test consisted of four sheets, each containing one of the following questions: (a) "What could be done to make a pump more effective, that is, to move more liquid or gas more rapidly?"; (b) "What could be done to make a pump more reliable, that is, to make sure it would not fail?"; (c) "Suppose you push down and pull up the handle of a pump several times but nothing comes out. What could have gone wrong?"; and (d) "Why does air enter a pump? Why does air exit from a pump?". A problem-solving score, which was open-ended, was determined by tallying the total number of acceptable solutions on the four problems.

The computer-based materials consisted of three programs respectively presenting (a) three presentations of an animation and narration simultaneously (concurrent program), (b) three sequential presentations of an animation followed by a narration (A-N successive program), and (c) three sequential presentations of a narration followed by an animation (N-A successive program). Before each presentation, instructions on the screen directed the student to press any key to begin. The animation lasted approximately 30 s and showed a line diagram of a tire pump, including handle, rod, piston, cylinder, inlet valve, outlet valve, and hose. The animation was based on an illustration in the World Book Encyclopedia (1991); the narration showed (a) the movement of each part, as well as air entering the pump when the handle was pulled up, and (b) the movement of each part, as well as air exiting the pump when the handle was pushed down. The narration was based on an article on pumps in the World Book Encyclopedia (1991) and was spoken in a male voice. The narration lasted approximately 30 s and contained the following words: "When the handle is pulled up, the piston moves up, the inlet valve opens, the outlet valve closes, and air enters the lower part of the cylinder. When the handle is pushed down, the piston moves down, the inlet valve closes, the outlet valve opens, and air moves out through the hose." The programs were written in HyperCard (Claris, 1990), the animations were developed using MacroMind Director (MacroMind, 1989), and the narrations were developed using MacRecorder (Farallon Computing, 1989). Selected frames of the animation along with the corresponding narration are shown in Figure 4.

The apparatus consisted of three Macintosh IIci computer systems, which included 13-in. color monitors and 80-megabyte internal hard drives.

Procedure. Students were tested in groups of one to three per session; each student was randomly assigned to a treatment group and then spatial ability was determined. First, students completed a questionnaire that solicited information concerning their previous experience with mechanical devices. Students who indicated extensive mechanical experience, by obtaining a score of five or above, were not used in the study. Second, students had 3 min to take the mental rotation test and 3 min to take the paper folding test. Students who scored above the median on the composite score were classified as high in spatial ability, and those who scored below the median were classified as low in spatial ability. Third, students were seated in front of a Macintosh IIci computer system and received instruction in how a bicycle tire pump works. Concurrent students received three presentations of a 30-s animation of a pump that was coordinated with the narration (i.e., concurrent program). Successive students received three presentations of the sequence, animation followed by narration (i.e., successive program A-N) or three presentations of sequence of the narration followed by animation (i.e., successive program N-A). Concurrent students received no instruction; therefore, performance of this group served as a baseline for assessing the knowledge of college students on this topic. Fourth, students had 2.5 min per problem to write as many answers as possible for each of four problem-solving transfer questions.

Results and Discussion

Scoring. As in previous studies (Mayer & Anderson, 1992), we created a list of acceptable responses for each of the four problems. For example, acceptable answers for the first problem on how to improve the pump's effectiveness included increasing the size of the cylinder and pressing the handle harder; acceptable answers for the second question on how to improve the pump's reliability include using airtight seals and creating a back-up system; acceptable answers for the third question about diagnosing a pump's failure include looking for a hole in the cylinder or for a stuck valve; and acceptable answers for the fourth

respectively. An analysis of variance (ANOVA) revealed that the scores differed significantly, F(2, 80) = 55.37, p < .001, MS_e = 3.76; supplemental Tukey tests (α = .05) indicated that the means of the concurrent and successive groups did not differ from one another, but each was significantly greater than the mean of the control group. The mean retention scores (and standard deviations) for the concurrent, successive, and control groups in Experiment 2 were 5.9 (1.7), 6.2 (2.0), and 3.2 (1.6), respectively. An ANOVA revealed that the scores differed significantly, F(2, 91) = 28.42, p < .001, MS_e = 3.15; supplemental Tukey tests (α = .05) indicated that the means for the concurrent and successive groups did not differ from one another, but each was significantly greater than the mean of the control group. These results replicate previous findings (Mayer & Anderson, 1991, 1992) and are consistent with the predictions of a dual-coding theory of multimedia learning.
"When the handle is pulled up, the piston moves up, the inlet valve opens, the outlet valve closes, and air enters the lower part of the cylinder."

"When the handle is pushed down, the piston moves down, the inlet valve closes, the outlet valve opens, and air moves out through the hose."

Figure 4. Selected frames and sound track from animation and narration on how pumps work—Experiment 1. Adapted from The World Book Encyclopedia, Vol. 15, p. 900, 1994. ©1994 World Book, Inc. By permission of the publisher.
question concerning why air enters and exits from a pump include discussions of suction and compression, respectively. We determined the problem-solving scores for each student by tallying the number of acceptable solutions for each problem.

**Contiguity effect.** Figure 5 summarizes the problem solving scores for high and low visual learners in each treatment group. As in previous research and as predicted by dual-coding theory, there was a contiguity effect in which students who received concurrent presentation performed better on the problem-solving test than did students who received successive presentation or no instruction. This observation was confirmed by a 3 (condition) × 2 (spatial ability) SAS Type III sums of squares analysis of variance (ANOVA) on the problem-solving data for all students that revealed a main effect for treatment group, F(2, 80) = 3.95, p < .05, MSe = 6.77. Supplemental Tukey tests (α = .05) revealed that the concurrent group scored significantly higher than did the successive and control groups, which did not differ significantly from one another. The lack of difference between the successive and control groups indicates that the separated visual and verbal instruction provided to successive students was quite ineffective. This finding is consistent with the idea that successive and control students were equally unable to build referential connections. Thus, ineffective instruction (i.e., successive instruction) facilitated problem-solving transfer no more than no instruction.

**Contiguity effect for low- versus high-spatial ability learners.** Next, we explored whether the contiguity effect depended on the spatial ability of the learners. According to the enhancement view of spatial ability, low-spatial ability students must devote more cognitive resources to building visual representational connections during learning than do high-spatial ability students, and therefore they can allocate fewer cognitive resources to building referential connections during learning. Building representational and referential connections is needed for effective problem-solving performance and is enabled by concurrent rather than successive presentation. The enhancement view of spatial ability predicts a strong contiguity effect for high— but not for low—spatial ability learners. As predicted by this enhancement view, high—spatial ability students who received concurrent presentation of animation and narration generated approximately 50% more creative solutions on transfer problems than did high—spatial ability students who received successive presentation or no instruction; an ANOVA revealed significant differences among the groups, F(2, 35) = 4.75, p < .05, MSe = 7.81, and supplemental Tukey tests (α = .05) revealed that the concurrent group scored higher than did the other two groups and that the other two groups did not differ from one another. Also as predicted, low—spatial ability students who received concurrent presentation of animation and narration did not generate significantly more creative solutions on transfer problems than did students who received successive presentation or no presentation, F(2, 45) < 1, MSe = 5.95. As predicted by the enhancement view of spatial ability, these results yielded a pattern in which concurrent presentation significantly benefited high—spatial ability learners but not low—spatial ability learners. An ANOVA performed on the data for all students revealed a main effect for spatial ability, F(1, 80) = 4.94, p < .05, MSe = 6.77. However, the interaction failed to reach statistical significance, F(2, 80) = 2.50, p < .09.

**Experiment 2**

In Experiment 2, low-experience students who scored low or high in spatial ability learned about how the human respiratory system works. As in Experiment 1, students received visual and verbal presentations concurrently, successively, or not at all. In Experiment 1, the order of presentation for the successive group had no significant effect on test performance (see Footnote 2), so the successive presentation in Experiment 2 involved only narration followed by animation.

**Method**

**Participants and design.** The participants were 97 college students who lacked extensive prior knowledge about human anatomy, including 17 high—spatial ability students and 15 low—spatial ability students in the concurrent group, 15 high—spatial ability students and 18 low—spatial ability students in the successive group, and 26 control students. The participants were 97 college students who scored low or high in spatial ability. According to the enhancement view of spatial ability, high—spatial ability students who received concurrent presentation of animation and narration generated approximately 50% more creative solutions on transfer problems than did high—spatial ability students who received successive presentation or no instruction; an ANOVA revealed significant differences among the groups, F(2, 35) = 4.75, p < .05, MSe = 7.81, and supplemental Tukey tests (α = .05) revealed that the concurrent group scored higher than did the other two groups and that the other two groups did not differ from one another. Also as predicted, low—spatial ability students who received concurrent presentation of animation and narration did not generate significantly more creative solutions on transfer problems than did students who received successive presentation or no presentation, F(2, 45) < 1, MSe = 5.95. As predicted by the enhancement view of spatial ability, these results yielded a pattern in which concurrent presentation significantly benefited high—spatial ability learners but not low—spatial ability learners. An ANOVA performed on the data for all students revealed a main effect for spatial ability, F(1, 80) = 4.94, p < .05, MSe = 6.77. However, the interaction failed to reach statistical significance, F(2, 80) = 2.50, p < .09.

**Figure 5.** Mean number of solutions generated on problem-solving test for low- and high-spatial ability students in each group—Experiment 1. For the low-spatial ability students the means (and standard deviations) for the concurrent, successive, and control groups are 5.42 (2.54), 5.05 (2.46), and 5.00 (2.32), respectively; for the high-spatial ability students the corresponding figures are 8.70 (2.58), 6.10 (3.15), and 4.72 (1.60).

---

2 We combined the two successive subgroups because (a) our theoretical focus yielded a research plan that called for combining the successive subgroups, (b) in previous research the successive subgroups did not differ significantly from one another on problem-solving scores (Mayer & Anderson, 1992), and (c) the mean problem-solving scores of the two subgroups (mean for A–N = 5.5; mean for N–A = 5.8) did not differ significantly from one another in the present study, t(40) = .36, p > .5.
group, and 17 high-spatial ability students and 15 low-spatial ability students in the control group.

Materials and apparatus. The paper-and-pencil materials consisted of a student questionnaire, a problem-solving test, and two spatial ability tests. The questionnaire was a one-page checklist designed to measure each student's prior knowledge of the human respiratory system. It included questions concerning the student's high school and college courses in anatomy, extracurricular experience such as cardiopulmonary resuscitation training (CPR), and the student's self-rating of knowledge about the human body on a 5-point scale ranging from very low (1) to very high (5).

The problem-solving test consisted of seven sheets, each containing a question and space for a written answer. For example, three questions were (a) "Suppose you are a scientist trying to improve the human respiratory system. How could you get more oxygen into the bloodstream faster?" (b) "Not enough oxygen is getting to the brain, and a person is about to faint. What could be wrong with the respiratory system?" (c) "You are giving a lecture on how pollution hurts the human body. How might you substantiate your claim that not only the lungs are affected by polluted air?"
The two spatial ability tests were identical to those used in Experiment 1.

The computer-based materials consisted of two computer programs that presented an approximately 45-s animation and a 100-word narration of how the human respiratory system works, each of which was repeated three times. The total presentation lasted approximately 2 min for the concurrent program and 4 min for the successive program. The programs were based on text and figures presented in fifth-grade science textbooks (Cooper, Blackwood, Boescher, Giddings, & Carin, 1985; Mallinson, Mallinson, Smallwood, & Valentinio, 1985) and explained three phases in respiration—inhaling oxygenated air into the lungs, exchanging oxygen from the lungs and carbon dioxide from the bloodstream, and exhaling carbon-dioxide rich air from the lungs. The soundtrack, which was in a female voice, was either presented before the animation (successive program) or synchronized with it (concurrent program). Selected frames from the animation along with the soundtrack are presented in Figure 6. In the animation, the body parts were drawn with black lines, the oxygenated air was represented in pink, oxygenated blood was red, air containing carbon dioxide was represented in light blue, and blood containing carbon dioxide was colored dark blue. As in Experiment 1, the sound was created using MacRecorder (Farallon Computing, 1989), and the animations were created using Macromind Director (Macromind, 1989). The apparatus consisted of the same computer systems that we used in Experiment 1.

Procedure. Students were tested in groups of 1 to 3, and each student was randomly assigned to a treatment group. First, students completed a questionnaire that solicited information concerning their previous experience with human anatomy. Students who indicated extensive knowledge of human anatomy were not used in the study. Second, students were seated in front of a Macintosh Iici computer system and received instruction in how the human respiratory system works on the basis of the description of the lungs in several elementary school science textbooks. Concurrent students received three presentations of a 45-s animation of the respiratory system (including mouth, nose, bronchial tubes, lungs, air sacs, bloodstream, and diaphragm) that was coordinated with the narration:

There are three phases in respiration—inhaling, exchanging, and exhaling. During inhaling, the diaphragm moves down creating more space for the lungs, air enters through the nose or mouth, moves down through the throat and bronchial tubes to tiny air sacs in the lungs. During exchange, oxygen moves from the air sacs to the bloodstream running nearby, and carbon-dioxide moves from the bloodstream to the air sacs. During exhaling, the diaphragm moves up, creating less room for the lungs, air travels through the bronchial tubes and throat to the nose and mouth, where it leaves the body.

Successive students received three presentations of the sequence—narration followed by animation. Control students received no instruction. Third, students had 2 min per problem to write as many answers as possible to each of the problem-solving transfer questions. Fourth, all students completed the same spatial tests that the students in Experiment 1 completed.

Results and Discussion

Scoring. The spatial-ability and problem-solving tests were scored like they were in Experiment 1. Only data from the three questions listed in the materials section were used for the problem-solving test, because these questions were most similar to those used in Experiment 1 (i.e., questions about how to modify the system and how to troubleshoot the system) and because the other questions proved to be difficult to score.

Contiguity effect. In Figure 7 we summarize the problem-solving scores for high- and low-visual learners in each treatment group. As in Experiment 1 and as predicted by dual-coding theory, there was a contiguity effect in which students who received the simultaneous presentation performed better on the problem-solving test than did students who received the successive presentation or no instruction. This observation was confirmed by a 3 (condition) X 2 (spatial ability) ANOVA on the problem-solving data for all students that revealed a main effect for treatment group, F(2, 91) = 22.61, p < .001, MSe = 2.54; supplemental Tukey tests (α = .05) revealed that the concurrent group scored significantly higher than did the successive group, which scored significantly higher than the control group.

Contiguity effect for low— versus high—spatial ability learners. As in Experiment 1, our focus in this study was on whether the contiguity effect is dependent on the spatial ability of the learners. Such a finding is in line with the predictions of the enhancement view of spatial ability. As predicted, high-spatial ability students who received the concurrent presentation of animation and narration generated approximately 50% more creative solutions on transfer problems than did high-spatial ability students who received successive presentation or no instruction; an ANOVA conducted for high-spatial ability students revealed significant differences among the groups, F(2, 46) = 14.15, p < .001, MSe = 43.75, and supplemental Tukey tests (α = .05) revealed that the concurrent group scored higher than did both the successive group and the control group, which did not differ from one another. Also consis-

3 In Experiment 2, half the students received animations that contained written labels for the major parts of the system, and half of the students received animations that did not. However, as this treatment produced no significant main effects or interactions, we have not described it in this article.
There are three phases in respiration—inhaling, exchanging, and exhaling.

Inhaling

1. During inhaling, the diaphragm moves down creating more space for the lungs, air enters through the nose or mouth, moves down through the throat and bronchial tubes to tiny air sacs in the lungs.

Exchanging

3. During exchange, oxygen moves from the air sacs to the bloodstream running nearby, and carbon-dioxide moves from the bloodstream to the air sacs.

Exhaling

5. During exhaling, the diaphragm moves up, creating less room for the lungs, air travels through the bronchial tubes and throat to the nose and mouth, where it leaves the body.

Figure 6. Selected frames and sound track from animation and narration of how the human respiratory system works—Experiment 2.
tent with predictions, low-spatial ability students who received concurrent presentation of animation and narration generated approximately the same number of creative solutions on transfer problems as did students who received successive presentation. An ANOVA conducted for low-spatial ability students revealed significant differences among the groups, \( F(2, 45) = 9.14, p < .001, MS_e = 23.33 \), and supplemental Tukey tests (\( \alpha = .05 \)) revealed that the two treatment groups each performed significantly better than did the control group, but they did not differ from one another. These results partially replicate the pattern in Experiment 1 in which concurrent presentation benefited high-spatial ability learners more than low-spatial ability learners, and they provide replicatory support for the enhancement view of spatial ability. Consistent with this observed pattern, an ANOVA performed on the data for all students revealed a statistically significant interaction between spatial ability and treatment, \( F(2, 91) = 3.19, p < .05, MS_e = 2.54 \). Although students with high-spatial ability scored higher than did students with low-spatial ability in both experiments, the main effect for spatial ability failed to reach statistical significance in Experiment 2, \( F(1, 91) = 1.96, p > .10 \). However, our theoretical focus in both experiments was on whether the effects of instructional methods depended on spatial ability; our focus was not on the main effect of ability.

**Figure 7.** Mean number of solutions generated on problem-solving test for low- and high-spatial ability students in each group—Experiment 2. For the low-spatial ability students the means (and standard deviations) for the concurrent, successive, and control groups are 4.07 (1.58), 4.06 (1.92), and 1.93 (1.10), respectively; for the high-spatial-ability students the corresponding figures are 5.53 (2.04), 3.53 (1.25), and 2.45 (1.32).

**General Discussion**

First, these results provide a solid replication of the contiguity effect: Inexperienced students were better able to transfer what they had learned about a scientific system when visual and verbal explanations were presented concurrently than when visual and verbal explanations were separated. The theoretical explanation for this finding, derived from a dual-coding theory of multimedia learning, is that concurrent presentation of verbal and visual descriptions of a system increases the likelihood that students will be able to build connections between their mental representations of visually and verbally presented explanations.

Furthermore, this dual-coding theory predicts that successive presentation makes it more difficult for learners to form referential connections. In some cases, the difficulty may be so great that students who receive successive presentation do not differ from control students on tests of problem-solving transfer—as was found in Experiment 1. In other cases, the difficulty may be less strong so that successive students outperform control students on tests of problem-solving transfer—as was found in Experiment 2.

Second, although these experiments did not include a direct experimental test, the findings are consistent with previously reported results concerning the role of specific knowledge on the contiguity effect (Mayer & Gallini, 1990). Although we did not test whether the contiguity effect is strong for inexperienced but weak for experienced students, we did focus on the first half of this prediction. In the present experiments, all students lacked experience in the subject domain, so these are the kinds of learners who would be expected to benefit most from concurrent rather than successive instruction. As expected, these students displayed a strong contiguity effect. The theoretical explanation for this finding is that experienced learners are able to retrieve appropriate familiar knowledge (such as a visual representation of the system) from long-term memory as they read or listen to a verbal description of a passage, and thus, they can build connections between the retrieved system and the system described in words. Inexperienced learners are not able to build connections between the verbally described system and a familiar system because they lack relevant knowledge in long-term memory. If domain-specific knowledge compensates for a lack of coordination between visual and verbal instruction, then there should be a contiguity effect for low-knowledge but not for high-knowledge learners as similar to Table 1. The results of both studies tested and confirmed the first half of this prediction, so these results are consistent with the idea that domain-specific knowledge compensates for uncoordinated instruction rather than enhances coordinated instruction.

The new finding across both experiments concerns the role of spatial ability in the contiguity effect: The contiguity effect was strong for high-spatial ability students but not for low-spatial ability students. In contrast to domain-specific experience—which can compensate for uncoordinated presentation of visual and verbal explanations—spatial ability appears to enhance coordinated visual and verbal instruction. In these experiments, temporal coordination of animations and narrations is most useful for students who have adequate spatial ability. We interpret this interaction within the context of dual-coding theory. Low-spatial ability students must devote a large amount of cognitive effort to building a visual representation of the system, whereas for high-spatial ability students building a visual representation that is based on the animation is relatively effortless. Given that cognitive resources in working memory are limited,
high–spatial ability students are more able to allocate sufficient cognitive resources to building referential connections than are low–spatial ability students.

Overall, these results support a dual-coding theory of multimedia learning, which emphasizes the learner’s building of mental connections between visual and verbal representations. Our research suggests an interesting hypothesis concerning the ways in which domain-specific knowledge and spatial ability affect learning from words and pictures, namely that domain-specific knowledge compensates for uncoordinated instruction, whereas spatial ability enhances coordinated instruction. Furthermore, these results contribute to an evolving theory of the cognitive conditions required for meaningful learning from words and pictures, because they highlight the role of individual differences.

These results have practical and theoretical implications. On the practical side, it seems worthwhile for instructional designers to be alert to an instructional manipulation that has been shown to increase transfer performance by 50% across a number of studies. Our work signals the value of instructional materials that maximize the learner’s chances of building connections between words and pictures. Unfortunately, examples of a lack of coordination between animation and narration can be found in educational multimedia products currently in use (e.g., presentation of animation without simultaneous narration). On the theoretical side, it seems worthwhile to seriously consider the proposal that scientific learning is like structure mapping—in that a learner builds a representation that is based on the presented words and a representation that is based on the presented pictures (or knowledge from long-term memory), and the learner tries to understand the relation between the structures of the two representations. Meaningful learning involves more than building either a verbal or visual representation; the additional component is building referential connections between the two kinds of mental representations.

The following conditions that produced our effects may serve as situational limits to the generalizability of the findings. First, the material in these studies is expository; it explains how a system works. More research is needed to determine whether the same results would be obtained if the material had been a description of unrelated facts. Second, the instructional methods in these studies are designed to promote or inhibit the building of referential connections (or structure mapping between the two representations of the same system). If our methods had emphasized the building of representational connections, such as rote drill on the material, we would not have expected any differences between groups on problem-solving transfer. Third, the major dependent measure in our studies is problem-solving transfer; if we had focused solely on retention of the presented material, we would not have expected a significant effect. Consistent with this prediction, Footnote 1 shows that in both experiments the concurrent and successive groups did not differ significantly in overall amount recalled. Finally, the results must be interpreted in light of the fact that the instructional episode was brief—less than 5 min; more research is needed to determine whether similar results would be obtained if students had more time to explore the material.

Our version of dual-coding theory extends, rather than directly applies, Paivio’s (1986) dual-coding theory to the domain of problem solving. Although our predictions do not contradict Paivio’s theory, we have had to clarify assumptions about the cognitive conditions for problem-solving transfer. In our version of dual-coding theory for multimedia learning, structural understanding (as measured by problem-solving transfer) occurs only when referential connections between verbal and visual representations are made; Paivio’s dual-coding theory does not claim that this is always necessary or even possible and does not emphasize problem-solving transfer as a dependent measure. Additional research is required to pinpoint the relation between building referential connections during learning and performing on tests of problem-solving transfer. Another issue worthy of further study is that imagery may exert its influence not only during learning—as emphasized in this study—but also during testing. Postexperimental questionnaires about imagery usage during learning and testing could help clarify this issue (McIntosh, 1986).

In conclusion, our study begins to answer a question that we adapted from Larkin and Simon (1987): “For whom is a picture worth a thousand words?” First, students who possess domain-specific knowledge may not need visual aids with text because they can generate their own familiar analogical representations as they read or listen to an explanation of a system. Therefore, mainly inexperienced students—such as those in our studies—benefit from pictures being coordinated with words. Second, students who possess high levels of spatial ability are more likely than low–spatial ability students to be able to build mental connections between visually based and verbally based representations. Therefore, mainly high–spatial ability students benefit from pictures being coordinated with words. Our results, consistent with the dual-coding theory, suggest that low-experience, high–spatial ability students are the most likely to benefit from instruction that carefully synchronizes the presentation of verbal and visual forms of scientific explanation. Researchers need to examine more fully the role that individual differences might play in multimedia learning.

References


MULTIMEDIA LEARNING

By Arle A. Paivio

Accepted January 5, 1994 •

Received April 8, 1993

Revision received December 20, 1993

Accepted January 5, 1994

Published in volume 38, 1993, pages 396-456.


Received April 8, 1993

Revision received December 20, 1993

Accepted January 5, 1994