From: Proceedings of IEEE MULTIMEDIA systems 1999, (pp. 720-725). Florence, Italy. IEEE Press

Designing Multimedia Presentations with Animation: What Does the Research Say?

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

ABSTRACT
How can we help students to understand scientific explanations of cause-and-effect systems, such as how lightning storms develop? One promising approach involves multimedia presentation of explanations in visual and verbal formats, such as presenting a computer-generated animation synchronized with narration or on-screen text. In a review of four studies, we found evidence that presenting a verbal explanation of how a system works with an animation does not insure that students will understand the explanation unless research-based cognitive principles are applied to the design.

KEYWORDS
Human Computer Interaction: Multimodal Interaction.
Multimedia Applications: Educational Applications.
Multimedia Tools: Animation and Computer Graphics.

Date submitted: October 15, 1998

Send correspondence to: Roxana Moreno or Richard E. Mayer
Department of Psychology
University of California
Santa Barbara, CA 93106
Electronic mail address: moreno@psych.ucsb.edu or
mayer@psych.ucsb.edu

Designing Multimedia Presentations with Animation: What Does the Research Say?

The purpose of this paper is to propose a set of instructional design principles, as derived from a review of recent empirical studies on multimedia learning with animations. In all experiments, students viewed a computer-generated animation depicting the process of lightning formation which included cool air becoming heated and condensing to form a cloud, the rising of the cloud beyond the freezing level and forming crystals, the movement of updrafts and downdrafts, the building of electrical charges within the cloud, the division of positive and negative charges, a negative stepped leader traveling to the ground and meeting a positive stepped leader from the ground, the negative charges following the path to the ground, and the positive charges following the path towards the cloud. Different elements from the presentation were varied in order to analyze the effect that such variation had in students' performance. To assess students' learning of the material, they were asked to write explanations of how lightning forms (retention test), to give names for parts of an illustration (matching test), and to apply what they have learned to solve new problems (transfer test). Based on the results of each experiment, a different design principle will be proposed.

In defining multimedia learning it is useful to distinguish among *media*, *mode* and *modality*. Media refers to the system used to present instruction, such as a book-based medium or a computer. Mode refers to the format used to represent the lesson, such as words versus pictures. Modality refers to the information processing channel used by the learner to process the information, such as auditory versus visual (Mayer, 1997). Of particular interest for the present review is the study of how specific combinations of modes and modalities may affect students' learning of scientific explanations, such as when we combine visual-verbal material (i.e., text) or auditory-verbal material (i.e., narration) with visual-non-verbal materials (i.e., animations). People can integrate multiple stimulus into one meaningful experience. For example, they can associate the sound of thunder to the visual image of lightning in the sky. But when the process of lightning formation is to be taught with a computer, the instructional designer is faced with the need to choose between several alternative media in order to promote meaningful learning (Mayer & Moreno, 1998). Within the visual modality for example, the process of lightning may be shown by a static diagram, an animation, or a video, and it may be described by text (Scaife & Rogers, 1996). Within the auditory modality for example, the process of lightning may be accompanied by sound effects or background music and it may be described by a narration. What is the optimal design and why?

Issue 1: Combining Verbal and Non-Verbal Presentations

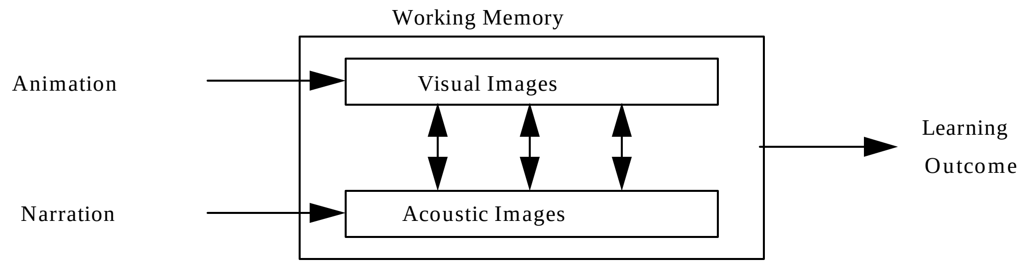
How should verbal information be presented to students to enhance learning from animations: auditorily as speech or visually as on-screen text? In order to answer this question, Mayer and Moreno (1998) asked students to view an animation depicting the process of lightning either along with concurrent narration (Group AN) or along with concurrent on-screen text (Group AT).

Our goal was to test a dual-processing theory of multimedia learning based on the following assumptions: (a) working memory includes an auditory working memory and a visual working memory, analogous to the phonological loop and visuo-spatial sketch pad, respectively, in Baddeley's (1986, 1992) theory of working memory; (b) each working memory store has a limited capacity, consistent with Sweller's (1988, 1989; Chandler & Sweller, 1992; Sweller, Chandler, Tierney & Cooper, 1990) cognitive load theory; (c) meaningful learning occurs when a learner retains relevant information in each store, organizes the information in each store into a coherent representation, and makes connections between corresponding representations in each store, analogous to the cognitive processes of selecting, organizing, and integrating in Mayer's generative theory of multimedia learning (1997; Mayer, Steinhoff, Bower & Mars, 1995); and (d) connections can be made only if corresponding pictorial and verbal information is in working memory at the same time, corresponding to referential connections in Paivio's (1986; Clark & Paivio, 1991) dual-coding theory.

Congruent with this dual-processing theory of multimedia learning, visually-presented information is processed--at least initially--in visual working memory whereas auditorily-presented information is processed--at least initially--in auditory working memory. For example, in reading text, the words may initially be represented in visual working memory and then be translated into sounds in auditory working memory. As shown in Figure 1, in the AN treatment, students represent the animation in visual working memory (such as the image of negative signs moving to the bottom of a cloud) and represent the corresponding narration in auditory working memory (such as the statement, "negative ions fall to the bottom of the cloud"). Because they can hold corresponding pictorial and verbal representations in working memory at the same time, students in group AN are better able to build referential connections between them, such as seeing that the image of negative signs moving to the bottom of the clouds corresponds to the words describing the fall of negative ions.

In the AT treatment, students try to represent both the animation and the on-screen text in visual working memory. Although some of the visually-represented text eventually may be translated into an acoustic modality for auditory working memory, visual working memory is likely to become overloaded. Students in group AT must process all incoming information--at least initially--through their visual working memory. Given the limited resources students have for visual information processing, using a visual modality to present both pictorial and verbal information can create an overload situation for the learner. If students pay full attention to on-line text they may miss some of the crucial images in the animation, but if they pay full attention to the animation they may miss some of the on-line text. Because they may not be able to hold corresponding pictorial and verbal representations in working memory at the same time, students in group AT are less able to build connections between these representations.

For Group AN, the incoming animation and narration initially are held in different working memory spaces.



For Group AT, the incoming animation and text initially are held in the same memory space.

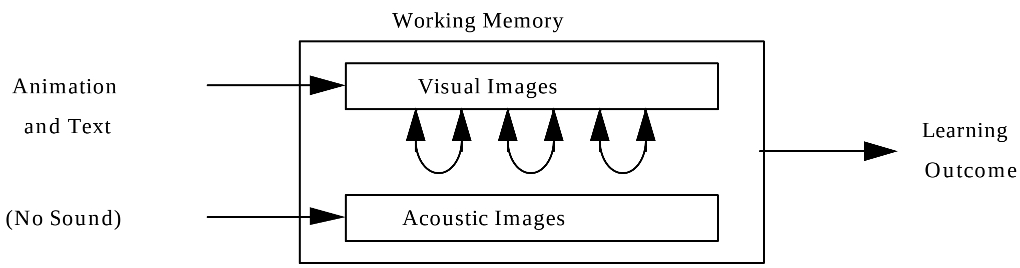


Figure 1. A dual-processing model of multimedia learning. (From Mayer & Moreno, 1998)

Therefore, dual-processing theory predicts that students in group AT perform more poorly than students in group AN on retaining the steps in the cause-and-effect chain (retention test), on being able match pictures and names of parts (matching test), and on being able to use what they have learned to solve problems (transfer test). The predictions are based on the idea that AT students may not have encoded as much of the verbal material as AN students, may not have been able to build as many referential connections between corresponding pictorial and verbal information as AN students, and may not have been able to construct a coherent mental model of the system as well as AN students.

Method

The participants were 78 college students who lacked knowledge of meteorology. They first viewed the animation with either concurrent narration in a male voice describing 8 major steps in lightning formation (Group AN) or concurrent on-screen text involving the same words and presentation timing (Group AT). Then, all students took the retention, transfer and matching tests.

Results

Figure 2 shows the proportion of correct answers on the retention, matching and transfer tests for the AN and AT groups.

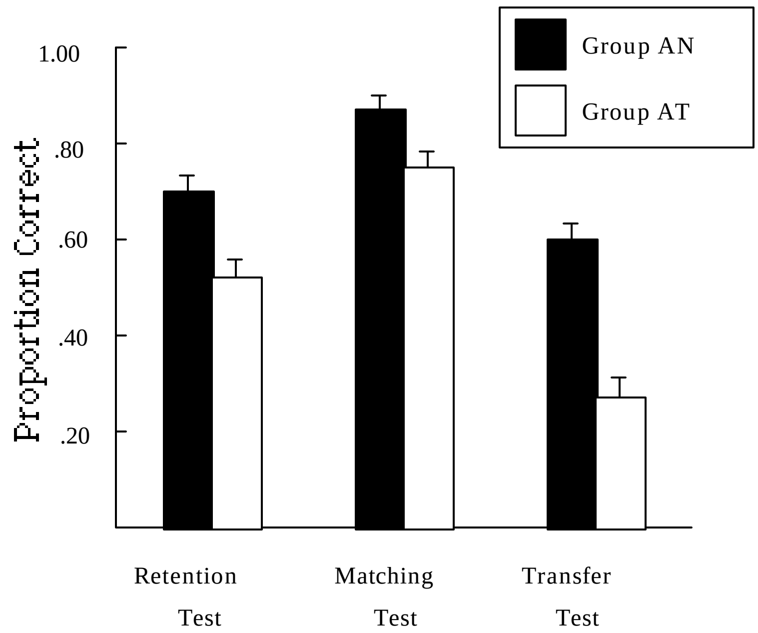


Figure 2. Proportion correct on retention, matching and transfer tests.
(From Mayer & Moreno, 1998)

Group AN recalled significantly ($p < .001$) more, correctly matched significantly ($p < .01$) more elements on diagrams, and generated significantly ($p < .001$) more correct solutions than Group AT. These results are consistent with the predictions of the dual-processing hypothesis and allow us to infer the first instructional design principle, called the split-attention principle by the cognitive load theory (Chandler & Sweller, 1992; Mousavi, Low, & Sweller, 1995).

Split-Attention Principle: Students learn better when the instructional material does not require them to split their attention between multiple sources of mutually referring information.

Issue 2: The Role of Modality

Why do students learn better when verbal information is presented auditorily as speech rather than visually as on-screen text? In Experiment 1, Mayer and Moreno (1998) showed that students who learn with concurrent narration and animations outperform those who learn with concurrent on-screen text and animations. However, 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 non-verbal) in working memory before attending to the other source. Therefore, the narration group might have had the advantage of being able to attend to both sources simultaneously, and the superior performance might disappear by using sequential multimedia presentations, where verbal and non-verbal materials are presented one after the other. The purpose of Experiment 2 was to test if the advantage of narration over on-screen text resides in a modality principle. If this is the case, then the advantage for auditory-visual presentations should not disappear when they are made sequential.

Method

The participants were 137 college students who lacked knowledge of meteorology. They first viewed the animation in one of the following 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). Similarly to Experiment 1, after viewing the animation, all students took the retention, transfer and matching tests.

Results

Figure 3 shows the proportion of correct answers on the retention, transfer and matching tests for the NN, AN, NA, AT, TA and TT groups.

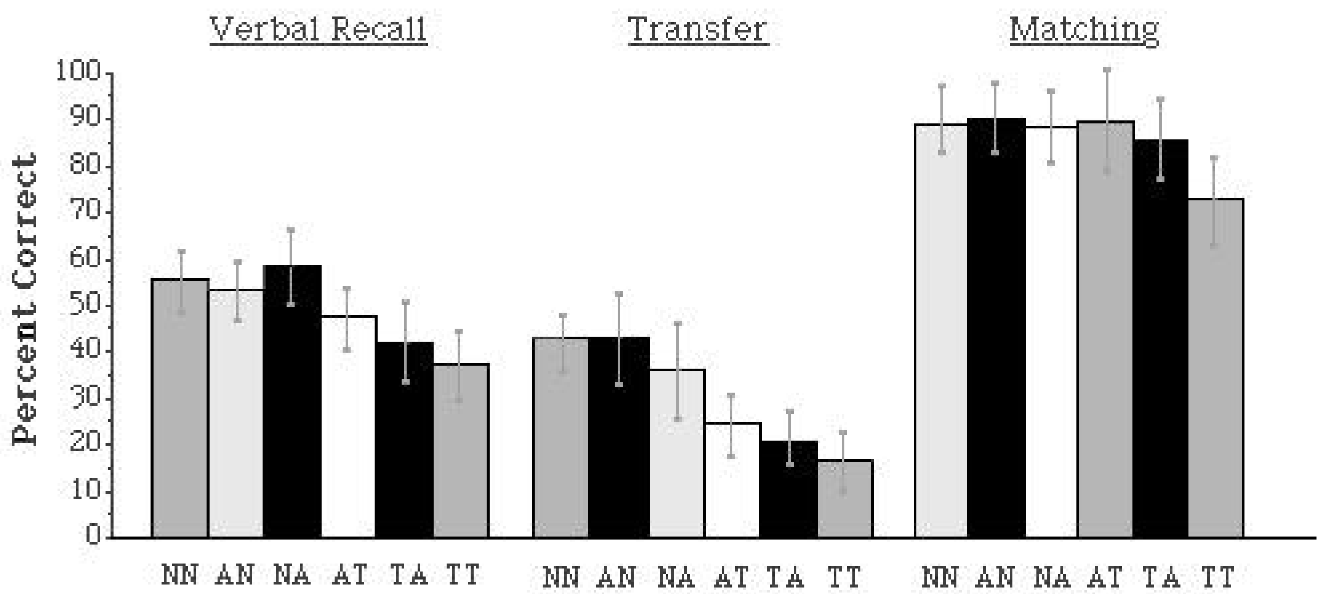


Figure 3. Proportion correct on retention, transfer and matching tests.

The text groups (TT, AT, and TA) scored significantly lower than the narration groups (NN, AN, and NA) in verbal recall ($p < .001$), problem solving transfer ($p < .001$), and matching ($p < .005$). These results reflect a modality effect.

Within each modality group, the simultaneous and sequential groups only showed a significant difference in their performance for matching tests ($p < .05$). This finding might be interpreted as an example of split-attention, where presenting two competing visual materials simultaneously has negative effects on the matching of verbal and visual materials in a multimedia lesson.

The results from Experiment 2 are consistent with prior studies on text and diagrams (Mousavi, Low, & Sweller, 1995), and allow us to infer a second instructional design principle--the Modality Principle.

Modality Principle: *Students learn better when the verbal information is presented auditorily as speech rather than visually as on-screen text both for concurrent and sequential presentations.*

Issue 3: The Role of Spatial Contiguity

How does the physical integration of on-screen text and visual materials affect students' learning? Mayer and Moreno's (1998) study showed that students who learn with concurrent narration and animations outperform those who learn with concurrent on-screen text and animations. These results might be caused by two different causes. First, and congruent with Experiment 2, students might be experiencing a modality effect, where information is processed more efficiently from two separate auditory and visual streams rather than from a sole visual stream. Second, students might be missing part of the visual information while they are reading the on-screen text (or vice versa). This second effect is similar to effects found in prior studies on text and pictures or diagrams. 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 ten studies concerning whether multimedia instruction is effective, Mayer (1997) concluded that there was a consistent evidence for what was called a spatial-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; Sweller & Chandler, 1994; Sweller, Chandler, Tierney & Cooper, 1990; Paas & Van Merrienboer, 1994). The goal of Experiment 3 was to distinguish between spatial-contiguity and modality effects in multimedia learning with animations. In order 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) while 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 while any differences in the performance of the narration group (N) respect to the text groups (IT and ST) can be interpreted exclusively in terms of modality.

<u>Method</u>

The participants were 132 college students who lacked knowledge of meteorology. They first viewed the animation in one of the following three conditions: N, ST or IT. Similar to Experiment 1, one group of students listened to concurrent narration describing each of the major events in words spoken at a slow rate by a male voice (Group N). The ST and IT versions instead, included concurrent on-screen text 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. As in Experiments 1 and 2, after viewing the animation, all students took the retention, transfer and matching tests.

<u>Results</u>

Figure 4 shows the proportion of correct answers on the retention, transfer and matching tests for the N, IT and ST groups.

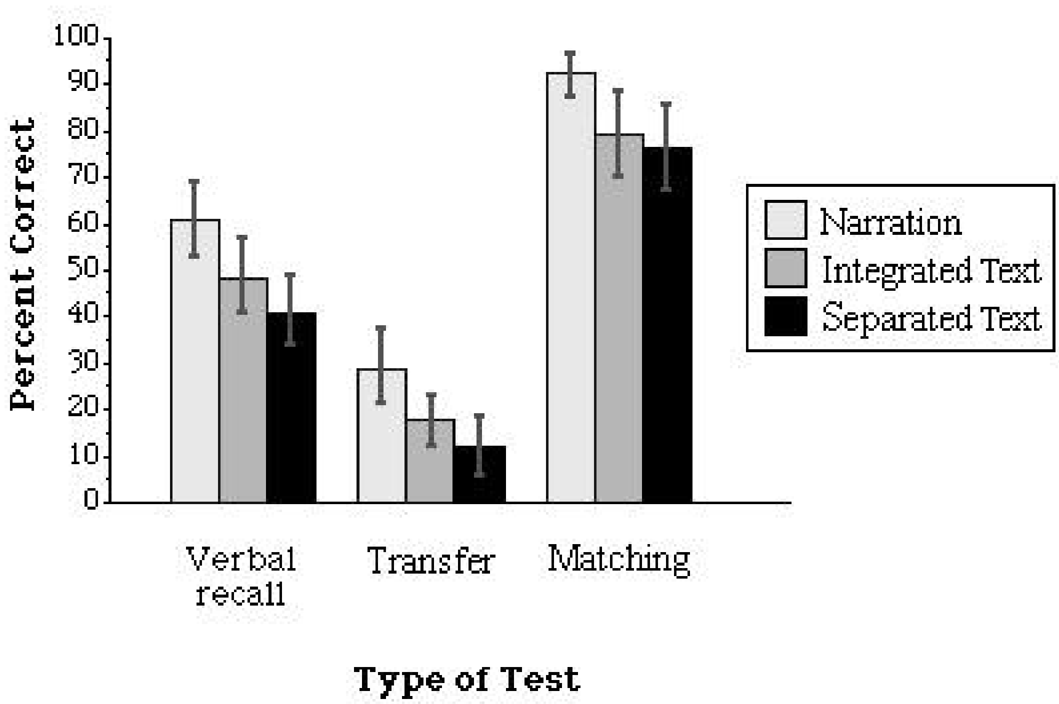


Figure 4. Proportion correct on retention, transfer and matching tests.

The N group scored significantly higher than the IT and ST groups in the verbal recall, and transfer tests, with the IT group scoring significantly higher than the ST group ($p < .001$). These results are evidence for both, a modality and a spatial-contiguity effect in recall and transfer. For matching tests, students in group N group scored significantly higher than the IT and ST, which did not differ from each other($p < .05$). These results are evidence for a modality effect in matching but not for a spatial-contiguity effect. 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.

Overall, it is possible to infer a third instructional design principle--the Spatial Contiguity Principle, which extends prior findings on text and diagrams to multimedia learning with animations

Spatial Contiguity Principle: *Students learn better when on-screen text and visual materials are physically integrated rather than separated.*

Issue 4: The Role of Temporal Contiguity

How synchronized in time do the verbal and visual materials have to be for students to learn better from animations? The results from Experiment 2 failed to find performance differences between the simultaneous and sequential narration groups, which seems to contradict the temporal contiguity effect found in some of the 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, Low and Sweller obtained similar results than those reported in Experiment 2. In a study where students studied geometry worked 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, Low & Sweller, 1995). How can we reconcile these findings?

Congruent with a dual-processing model of working memory, constructivist learning is fostered when the learner is able to hold a visual representation in visual working memory and a corresponding verbal representation in verbal working memory at the same time. The model implicates working memory load as a major impediment to learning. Although all learners may receive identical animation and narration in a multimedia environment, the amount of information that they are forced hold in working memory at one time might affect performance. For example, presenting the whole animation preceded or followed by the whole narration successively (which we will call large bites), can overload working memory such that it is not possible to hold all of the narrative in working memory until the animation is presented (or vice versa). However, if we present one chunk of animation--depicting only a short sequence--preceded or followed by one corresponding chunk of narration and so on (which we will call small bites), and if the size of the chunk does not exceed working memory capacity, then the learner should be able to make connections between the corresponding words and pictures--in the same way as when animation and narration are presented concurrently.

In the Mayer and Anderson (1991, 1992) studies, large bites 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, Low and Sweller and Experiment 2, the successive presentations of the animation and narration were small bites, only a line or two at a time, and thus unlikely to overload working memory.

The purpose of Experiment 4 is to reconcile the above findings by varying the cognitive load of the learners. In order to do so, the performance of three groups of students was compared. One group of students saw a presentation with concurrent animation and narration (concurrent group), a second group of students either viewed the whole animation followed or preceded by the whole explanatory narration (large bites group) and a third group of students either viewed small chunks of animation followed or preceded by small chunks of explanatory narration (small bites group).

Method

The participants were 60 college students who lacked knowledge of meteorology. They first viewed the animation in one of the following conditions: concurrent narration and animation, large bites of narration and animation or small bites of narration and animation. As in Experiments 1, 2 and 3, after viewing the animation, all students took the retention, transfer and matching tests.

Results

Figure 5 shows the number of correct answers (over a total possible score of 8) on the retention, transfer and matching tests for the narration, small bites and large bites groups. The large bites group scored significantly lower than the concurrent and small bites groups in the verbal recall, matching and transfer tests, ($p = .002$, $p < .0001$, and $p < .0001$, respectively). The concurrent and small bites groups did not differ from each other.

These results allow all prior studies in temporal contiguity to be reconciled if the length of the cycles in the sequential presentations are taken into account. Therefore, a fourth instructional design principle for multimedia learning with animations can be inferred. *Temporal Contiguity Principle: Students learn better when verbal and visual materials are temporally synchronized rather than separated in time.*

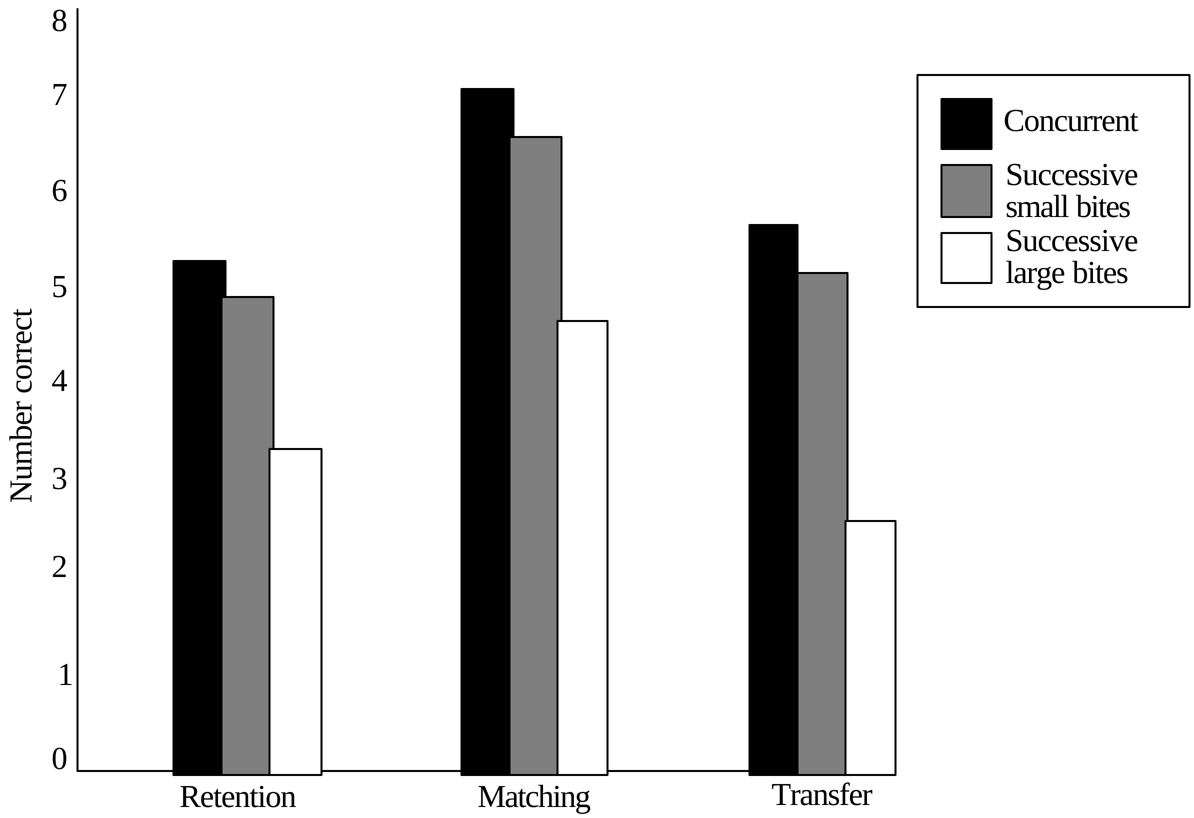


Figure 5. Number correct on retention, matching, and transfer tests.

Conclusion

Although multimedia learning offers very high potential educational opportunities, computer-based instructional materials are usually based on what current technology advances can do rather than on research-based principles of how students learn with technology. Multimedia environments allow students to work easily with verbal and non-verbal representations of complex systems. They also allow the use of different modalities to present the same information. The present review demonstrates that presenting a verbal explanation of how a system works with an animation does not insure that students will understand the explanation unless research-based principles are applied to the design.

Our first study showed that students learn better from designs that do not present mutually referring information simultaneously. The split-attention principle emphasizes the need to present animation with auditory speech rather than on-screen text. Presenting an animation with simultaneous on-screen text forces students to hold one source of the visual materials in working memory while attending to the other source, creating a high cognitive load.

In the second study, evidence was found for a modality principle, where students learn better if the verbal material is presented auditorily rather than visually. It showed that the advantage of narration presentations over on-screen text presentations does not disappear when both groups are forced to hold the information contained in one source of the materials before attending to the other. These results suggest not only that more information is likely to be held in both auditory and visual working memory rather than in just one but that the combination of auditory verbal materials with visual non-verbal materials may create deeper understanding than the combination of visual verbal and non-verbal materials.

The third study examined the contribution of spatial contiguity to multimedia learning. A spatial contiguity principle was derived from the finding that learning is impaired when on-screen text is spatially separated from the visual materials.

Our last study found evidence for a temporal contiguity principle, according to which students learn better from an animation the more synchronized in time it is with the corresponding explanatory narration.

In our studies, the materials and learners may have contributed to the results. First, we used visual and verbal materials that explained a cause-and-effect system. Had we focused on a different learning context such as descriptive passages that presented lists of facts, we might have obtained different results. Second, we tested only low-experience students. Based on past research, high experience subjects would have been less likely to exhibit the effects that split attention, spatial contiguity temporal contiguity and modality produced in our studies (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. Third, we did not examine individual differences such as spatial ability, working memory span or coordination ability. Additional research is needed to determine the role of individual differences in multimedia learning.

References

Baddeley, A.D. (1986). Working memory. Oxford, England: Oxford University Press.

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

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.

Clark, J.M. & Paivio, A. (1991). Dual coding theory and education. Educational Psychology Review, 3, 149-210.

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

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

Mayer, R.E. & Anderson, R.B. (1991). Animations need narrations: An experimental test of a 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., 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 and Development, 43, 31-43.

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.

Mousavi, S.Y., 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 representation: A dual coding approach. Oxford, England: Oxford University Press.

Scaife, M. & Rogers, Y. (1996). External cognition: How do graphical representations work. International Journal of Human- Computer Studies, 45, 185-213.

Sweller, J. (1988). Cognitive load during problem solving: Effects on learning. Cognitive Science, 12, 257-285.

Sweller, J. (1989). Cognitive technology: Some procedures for facilitating learning and problem solving in mathematics and science. Journal of Educational Psychology, 81, 457-466.

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 as a factor in structuring of technical material. Journal of Experimental Psychology: General, 119, 176-192.