Visual Presentations in Multimedia Learning: Conditions that Overload Visual Working Memory

Roxana Moreno and Richard E. Mayer

University of California, Santa Barbara
Psychology Department
Santa Barbara, CA 93106
(805) 893-2472
{Moreno, Mayer}@psych.ucsb.edu

Abstract. How should we design visual presentations to explain how a complex system works? 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 three 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. The first two studies revealed a split-attention effect, in which students learned better when the instructional material did not require them to split their attention between multiple visual sources of information. The third study, 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 two cognitive principles of multimedia learning.

1 Introduction

The purpose of this paper is to propose a set of instructional design principles for visual presentations, as derived from a review of recent empirical studies on multimedia learning. In all studies, students were presented with verbal and non-verbal visual information and their learning from the multimedia lesson was compared to that of students who were presented with identical graphics and animations but instead of viewing on-screen text, listened to a narration.

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 [5]. 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., graphics, video or animations).

In all studies, after viewing a multimedia presentation, students had to complete a series of tests aimed to assess their retention and learning. Participants were asked to write down as much of the material as they could remember (retention test), to give names for parts of the animation (matching test), and to apply what they have learned to solve new problems (transfer test). Based on the results of our studies, two design principles will be proposed: the split-attention principle, and the modality principle.

1.1 Issue 1: A Split-Attention Effect

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 [7] asked students to view an animation depicting a complex system (the process of lightning formation, or how a car's braking system works), 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 [1, 2] theory of working memory; (b) each working memory store has a limited capacity, consistent with Sweller's [3, 12, 13] 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 [5, 9]; 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 [4, 11] 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 and represent the corresponding narration in auditory working memory. 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.

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.

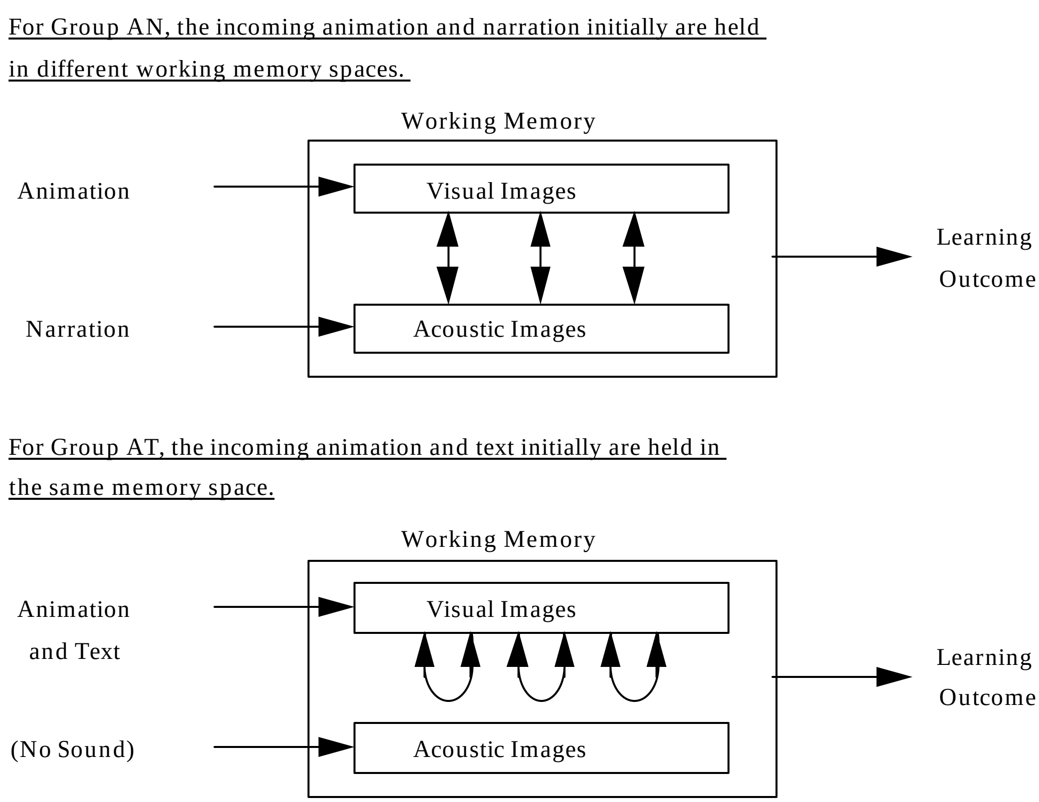

Fig. 1. A dual-processing model of multimedia learning. (From Mayer & Moreno, 1998)

Therefore, dual-processing theory predicts that students in group AT perform more poorly than students in group AN on retention, matching , and transfer tests. The predictions are based on the idea that AT students may not have encoded as much of the visual 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. Seventy eight college students who lacked knowledge of meteorology participated in the study of lightning formation, and 68 college students who had low knowledge of car mechanics participated in the study of a car's braking system. All participants first viewed the animation with either concurrent narration in a male voice describing the major steps in the respective domain (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. Figures 2 and 3 show the proportion of correct answers on the retention, matching and transfer tests for the AN and AT groups who viewed the lightning and car's braking system animation, respectively.

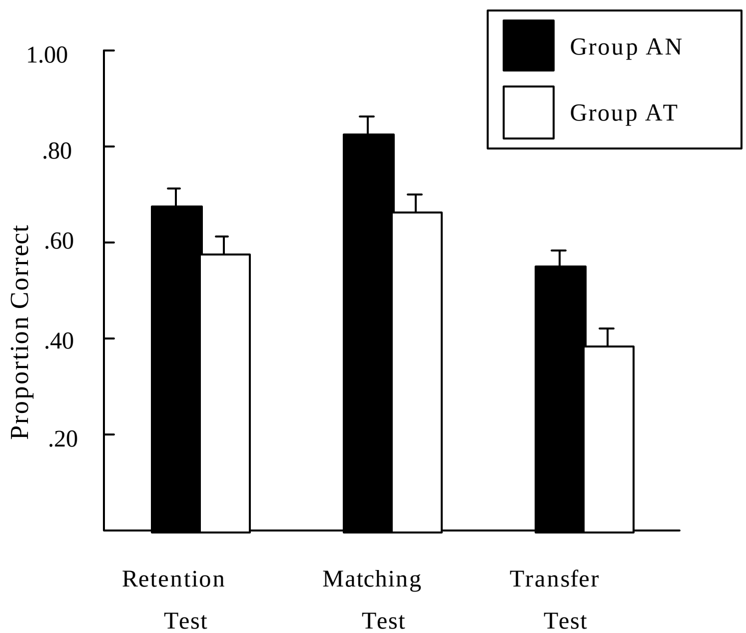

Fig. 2. Proportion correct on retention, matching and transfer tests--Lightning group. (From Mayer & Moreno, 1998)

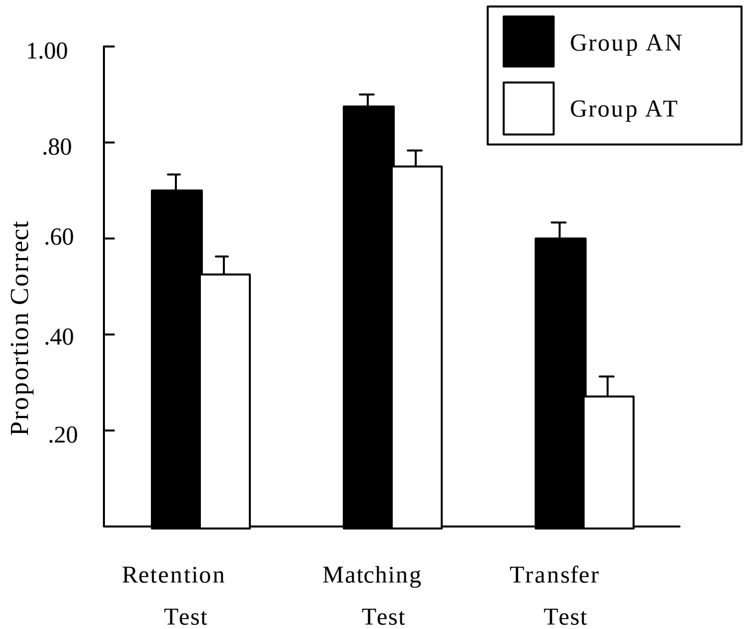

Fig. 3. Proportion correct on retention, matching and transfer tests--Car braking group. (From Mayer & Moreno, 1998)

In the lightning presentation, 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. Similarly, in the car braking presentation, group AN recalled significantly ($p< .05$) more, correctly matched significantly ($p < .05$) more elements on diagrams, and generated significantly ($p< .01$) 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 [3, 10]. *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.

1.2 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? Our first study showed that students who learn with concurrent narration and animation outperform those who learn with concurrent on-screen text and animation [7]. However, this type of concurrent multimedia presentations, 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 the second study 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, that is, when the graphics or animation are presented either before or following the narration or on-screen text. **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 our first two studies, 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). After viewing the animation, all students took retention, transfer and matching tests.

Results. Figure 4 shows the proportion of correct answers on the retention, transfer and matching tests for the NN, AN, NA, AT, TA and TT groups.

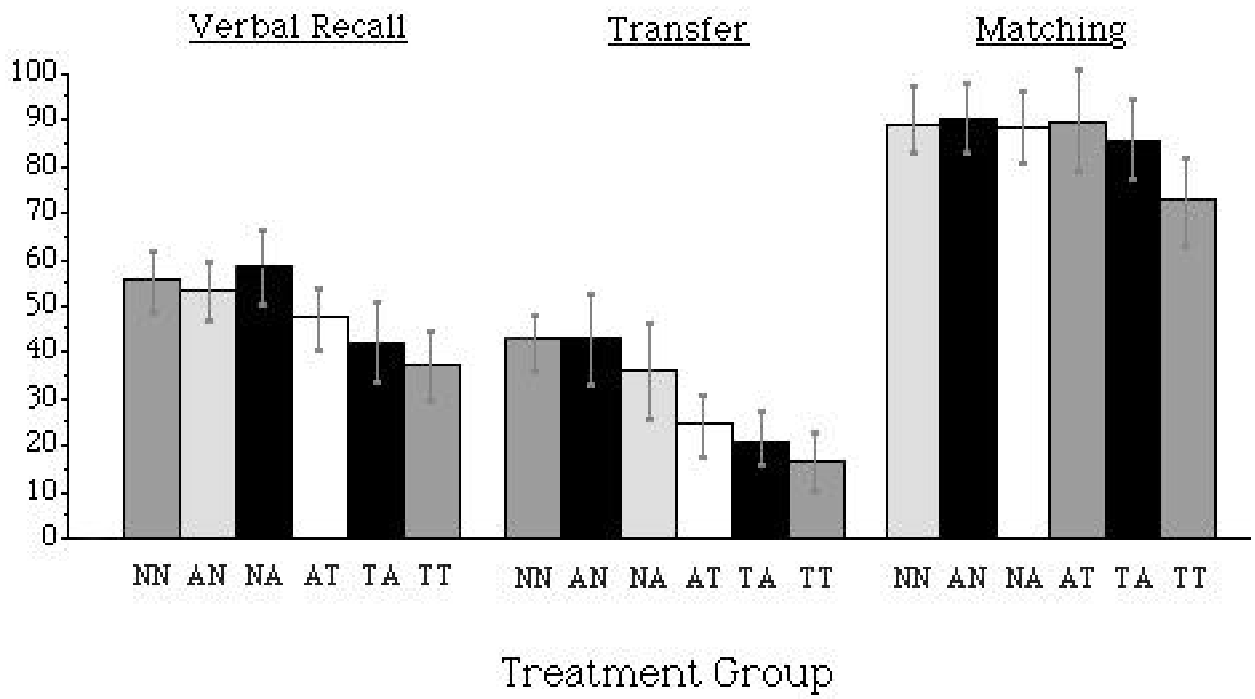

Fig. 4. Proportion correct on retention, transfer and matching tests.

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 association of verbal and visual materials in a multimedia presentation.

These results are consistent with prior studies on text and diagrams [10], 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.

2 General Discussion

2.1 Theoretical Implications

These results provide an important empirical test of a dual-processing theory of working memory within the domain of multimedia learning according to which students will learn better in multimedia environments when words and pictures are presented in separate modalities than in the same modality. When pictures and words are both presented visually (i.e., a split-attention situation), learners are able to select fewer pieces of relevant information because visual working memory is overloaded. When words and pictures are presented in separate modalities, visual working memory can be used to hold representations of pictures and auditory working memory can be used to hold representations of words. The robustness of these results was evident on two different domains (meteorology and mechanics) across three different studies.

Although multimedia learning offers very high potential educational opportunities by the presentation of rich visual information such as graphics, animation, and movies, 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 complex graphics, does not insure that students will remember or 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 simultaneous mutually-referring visual information. 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 even in sequential presentations. It showed that the advantage of narration presentations over on-screen text presentations does not disappear when both groups are forced to hold the infomation 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.

2.1 Practical Implications

This study calls attention to the need to broaden the goals of instructional designers of visual presentations. The design of multimedia presentations should be guided by the goal of presenting information that is relevant, and in a way that fosters active cognitive processing in the learner. Focusing solely on the first goal--presenting relevant information--can lead to presentations such as the one given to the AT groups in our studies, where visual working memory is likely to become overloaded. When working memory becomes overloaded, the opportunities for active cognitive processing are reduced. Focusing on both goals --presenting relevant information in ways that promote active learning--can lead to presentations such as the one given to the AN groups in our studies, where working memory is less likely to become overloaded.

An important consideration in the design of multimedia presentations is whether to accompany animations with auditorily-presented or visually-presented words. The most important practical implication of this study is that animations should be accompanied by narration rather than by on-screen text. This implication is particularly important in light of the increasing use of animations and on-screen text both in courseware and on the world wide web. These results cast serious doubts on the implicit assumption that the modality of words is irrelevant when designing multimedia presentations.

These results should not be taken as a blanket rejection of the use of text captions with graphics. To the contrary, in a series of studies on text and illustrations about how devices work carried out in our lab at Santa Barbara the results consistently have shown that students learn more productively when text is presented within corresponding illustrations rather than when text and illustrations are presented on separate pages [6, 5, 8, 9]. Similarly, in a series of studies on worked-out geometry problem examples Sweller and his colleagues have shown that students learn better when text explanations are presented on the sheet with geometry problems than separately [14, 15]. Overall, these studies provide ample evidence for the benefits of presenting short captions or text summaries with illustrations.

References

1. Baddeley, A.D.: Working memory. Oxford, England: Oxford University Press (1986)
2. Baddeley, A.: Working memory. Science, 255, 556-559 (1992)
3. Chandler, P. & Sweller, J.: The split-attention effect as a factor in the design of instruction. British Journal of Educational Psychology, 62, 233-246 (1992)
4. Clark, J. M. & Paivio, A.: Dual coding theory and education. Educational Psychology Review, 3, 149-210 (1991)
5. Mayer, R. E.: Multimedia learning: Are we asking the right questions? Educational Psychologist, 32, 1-19 (1997)
6. Mayer, R. E.: Systematic thinking fostered by illustrations in scientific text. Journal of Educational Psychology, 81, 240-246 (1989).

7. Mayer, R. E. & Moreno, R.: A split-attention effect in multimedia learning: Evidence for dual processing systems in working memory. Journal of Educational Psychology, 90, 312-320 (1998)
8. Mayer, R. E. & Gallini, J. K.: When is an illustration worth ten thousand words? Journal of Educational Psychology, 82, 715-726 (1990)
9. Mayer, R. E., Steinhoff, K., Bower, G. & Mars, R.: A generative theory of textbook design: Using annotated illustrations to foster meaningful learning of science text. Educational Technology Research and Development, 43, 31-43 (1995)
10. Mousavi, S. Y., Low, R., & Sweller, J.: Reducing cognitive load by mixing auditory and visual presentation modes. Journal of Educational Psychology, 87, 319-334 (1995)
11. Paivio, A,: Mental representation: A dual coding approach. Oxford, England: Oxford University Press (1986)
12. Sweller, J.: Cognitive technology: Some procedures for facilitating learning and problem solving in mathematics and science. Journal of Educational Psychology, 81, 457-466 (1989)
13. Sweller, J.: Cognitive load during problem solving: Effects on learning. Cognitive Science, 12, 257-285 (1988)
14. Tarmizi, R. & Sweller, J.: Guidance during mathematical problem solving. Journal of Educational Psychology, 80, 424-436 (1988)
15. Ward, M. & Sweller, J.: Structuring effective worked out examples. Cognition and Instruction, 7, 1-39(1990)