1 Introduction to Multimedia Learning

Richard E. Mayer
University of California, Santa Barbara

Abstract

Multimedia learning is learning from words and pictures. The rationale for studying multimedia learning is that people can learn more deeply from words and pictures than from words alone. A goal of research on multimedia learning is to understand how to design multimedia learning environments that promote meaningful learning. The research base concerning multimedia learning is reflected in the 34 chapters of this handbook. What is new in this second edition is a sharp increase in the research base, the addition of seven new principles of multimedia learning, a broadening of contexts for studying multimedia learning, a better delineation of boundary conditions for principles, and refinements of theories of multimedia learning. The approach taken in this handbook is learner-centered rather than technology-centered, views learning as a constructive process rather than solely as a process of adding new information to memory or strengthening associations, seeks to foster meaningful learning rather than rote learning, and favors appropriate cognitive activity during learning rather than behavioral activity per se.

Introduction

People can learn more deeply from words and pictures than from words alone. This seemingly simple proposition – which can be called the *multimedia learning hypothesis* – is the main focus of this second edition of *The Cambridge Handbook of Multimedia Learning*.[1] Each of the 34 chapters examines an aspect of the multimedia learning hypothesis. In particular, multimedia researchers are interested in how people learn from words and pictures and in how to design multimedia learning environments that promote learning. In this chapter, I provide a definition of multimedia learning, offer a rationale for multimedia learning, outline the research base for multimedia learning, summarize changes since the first edition, and draw distinctions between two approaches to multimedia design, three metaphors

[1] There may be some conditions in which words or pictures alone are better than words and pictures combined, such as the redundancy effect described by Sweller and Kalyuga in Chapter 10 and the expertise reversal effect described by Kalyuga in Chapter 24.

of multimedia learning, three kinds of multimedia learning outcomes, and two kinds of active learning.

What Is Multimedia Learning?

Table 1.1 summarizes definitions of multimedia, multimedia learning, and multimedia instruction.

Multimedia

The term *multimedia* conjures up a variety of meanings. You might think of watching a podcast on your smartphone or playing a strategy game on your tablet – that is, multimedia as a handheld experience. You might think of sitting in a room where images are presented on one or more screens and music or other sounds are presented via speakers – that is, multimedia as a "live" performance. Alternatively, you might think of sitting at a computer screen that presents graphics on the screen along with spoken words from the computer's speakers – that is, multimedia as an online lesson. Other possibilities include watching a video on a TV screen while listening to the corresponding words, music, and sounds or watching a PowerPoint presentation along with listening to the speaker's corresponding commentary. Low-tech examples of multimedia include a *chalk-and-talk* presentation, in which a speaker draws or writes on a blackboard (or uses an overhead projector) while presenting a lecture or a textbook lesson consisting of printed text and illustrations. In sum, most academic learning situations involve multimedia learning because students encounter words and graphics.

I define *multimedia* as presenting both words (such as spoken text or printed text) and pictures (such as illustrations, photos, animation, or video). By words, I mean that the material is presented in *verbal form*, such as printed text or spoken text. By pictures, I mean that the material is presented in *pictorial form*, such as static graphics, including illustrations, graphs, diagrams, maps, or photos, or dynamic graphics, including animation or video. This

Table 1.1. *Definitions*

Term	Definition
Multimedia	Presenting words (such as printed text or spoken text) and pictures (such as illustrations, photos, animation, or video)
Multimedia learning	Building mental representations from words and pictures
Multimedia instruction	Presenting words and pictures that are intended to promote learning

definition is broad enough to include all of the scenarios I described in the preceding paragraph – ranging from multimedia encyclopedias to online educational games to textbooks. For example, in a multimedia encyclopedia, words may be presented as narration and pictures may be presented as animation. In a textbook, words may be presented as printed text and pictures may be presented as illustrations. In an online educational game, on-screen characters may speak as they show you how to accomplish some task.

If multimedia involves presenting material in two or more forms, then an important issue concerns how to characterize a form of presentation. Three solutions to this problem are the delivery media view, the presentation modes view, and the sensory modalities view. According to the delivery media view, multimedia requires two or more delivery devices, such as a computer screen and amplified speakers or a projector and a lecturer's voice. According to the presentation modes view, multimedia requires verbal and pictorial representations, such as on-screen text and animation or printed text and illustrations. According to the sensory modalities view, multimedia requires auditory and visual senses, such as narration and animation or a lecture and slides.

I reject the delivery media view because it focuses on the technology rather than on the learner. Instead, I opt for the presentation modes view and, to some extent, the sensory modalities view. The presentation modes view allows for a clear definition of multimedia – presenting material in verbal and pictorial form – and is commonly used by multimedia researchers (Mayer, 2009). The presentation modes view is also the basis for Paivio's (1986, 2006) dual-code theory, as well as theories of multimedia learning presented in this handbook (Chapter 2, by Paas and Sweller; Chapter 3, by Mayer; Chapter 4, by Schnotz; and Chapter 5, by van Merriënboer and Kester). The sensory modalities view is also relevant because words can be presented as printed text (initially processed visually) or as spoken text (initially processed auditorily), whereas pictures are processed visually. In conclusion, as shown in Table 1.1, multimedia refers to using words and pictures.

Multimedia learning

Multimedia learning occurs when people build mental representations from words (such as spoken text or printed text) and pictures (such as illustrations, photos, animation, or video). As you can see from this definition, multimedia refers to the presentation of words and pictures, whereas multimedia learning refers to the learner's construction of knowledge from words and pictures. The process by which people build mental representations from words and pictures is the focus of Mayer's cognitive theory of multimedia learning (Mayer, 2009; see also Chapter 3), Sweller's cognitive load theory (Sweller, Ayres, & Kalyuga, 2011; see also Chapter 2), Schnotz's integrative model of text and picture comprehension (Schnotz & Bannert, 2003; see also Chapter 4), and, to some extent, van Merriënboer's four-component instructional design theory (van Merriënboer & Kirschner, 2007; see also Chapter 5).

Multimedia instruction

Multimedia instruction (or a multimedia learning environment) involves presenting words and pictures that are intended to promote learning. In short, multimedia instruction refers to designing multimedia learning environments in ways that help people build mental representations. The instructional design principles described in Parts II and III suggest ways of creating multimedia lessons intended to promote multimedia learning, and Parts IV and V offer examples of how the principles can be applied in a variety of advanced contexts ranging from educational games to intelligent tutoring systems.

What Is the Rationale for Multimedia Learning?

What is the value of adding pictures to words? Do students learn more deeply from words and pictures than from words alone? These questions are essential to the study of multimedia learning. For example, suppose I asked you to listen to a short explanation of how a bicycle tire pump works: "When the handle is pulled up, the piston moves up, the inlet valve opens,

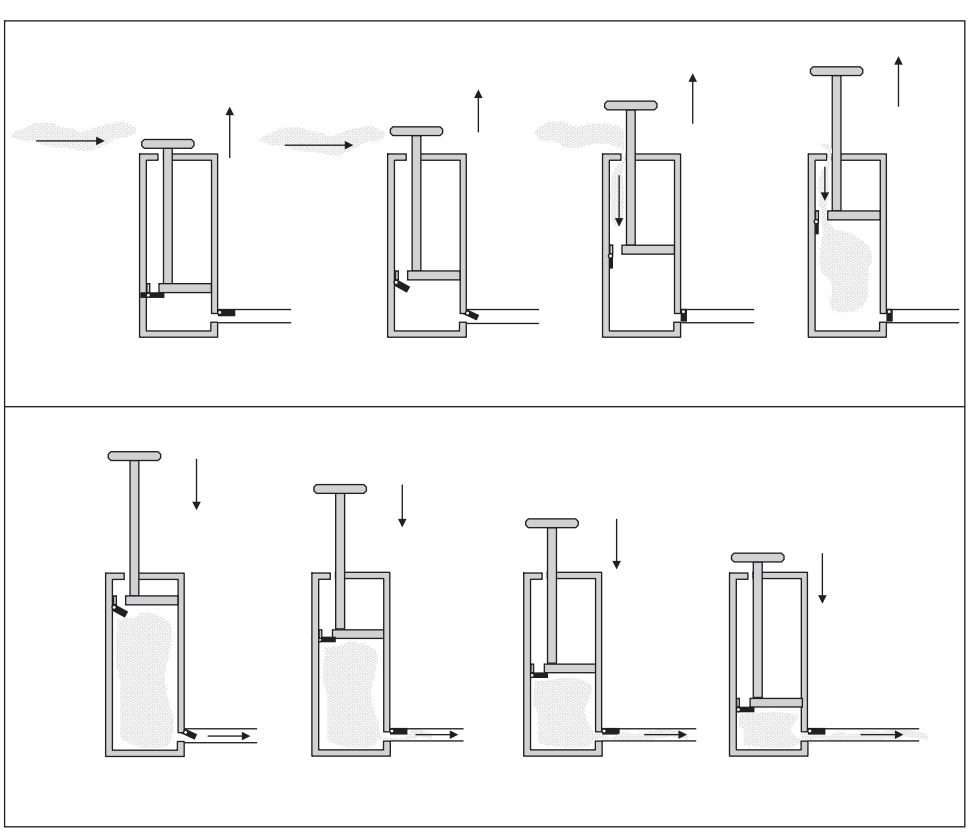

Figure 1.1. *Frames from a pumps animation.*

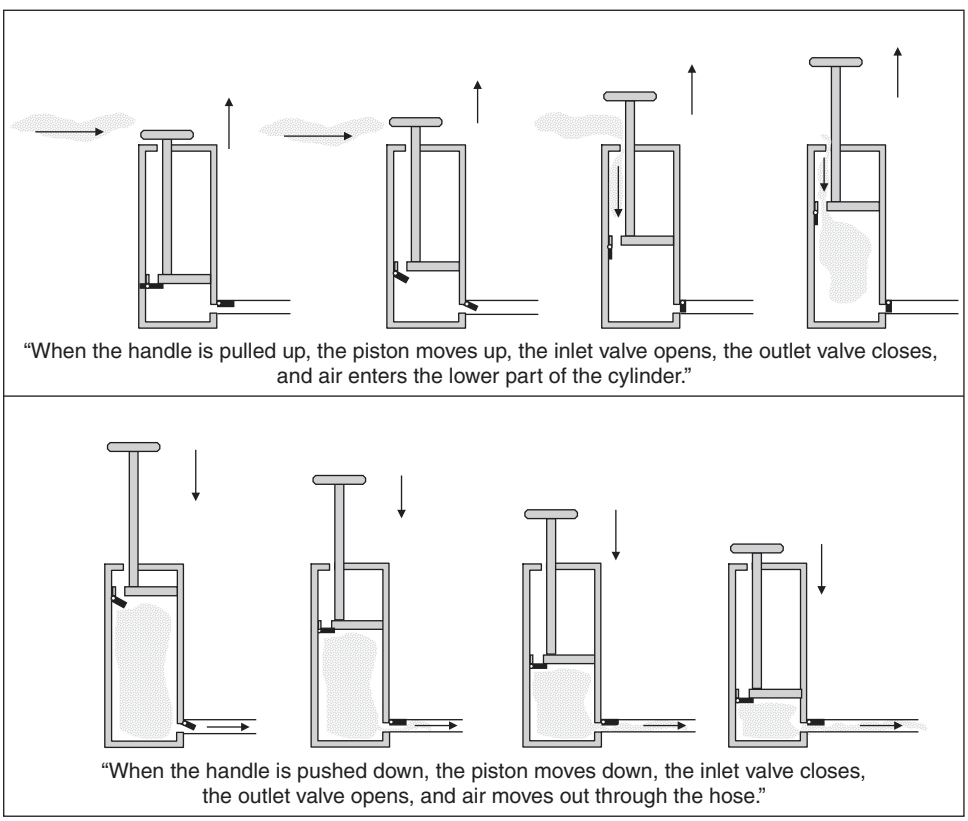

Figure 1.2. *Frames from a pumps animation with corresponding narration.*

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." Then I ask you to write down an explanation of how a bicycle tire pump works (i.e., retention test) and to write answers to problem-solving questions such as "Suppose you push down and pull up the handle of a pump several times but no air comes out. What could have gone wrong?" (i.e., transfer test). If you are like most of the students in our research studies (Mayer & Anderson, 1991, 1992), you remembered some of the words in the presentation (i.e., you did moderately well on retention) but you had difficulty using the material to answer problem-solving questions (i.e., you did poorly on transfer).

In contrast, suppose I showed you an animation of a bicycle tire pump that depicts the actions in the pump as the handle is pulled up and then as the handle is pushed down. Frames from the animation are shown in Figure 1.1. If you are like most students in our research studies (Mayer & Anderson, 1991, 1992), you would not do well on a retention test or on a transfer test.

Finally, consider the narrated animation summarized in Figure 1.2. In this situation, you hear the steps described in words as you see the steps depicted in the animation. When words and pictures are presented together

as in a narrated animation, students perform well both on retention and on transfer tests (Mayer & Anderson, 1991, 1992). In particular, when we focus on tests of problem-solving transfer – which are designed to measure a student's understanding of the presented material – students perform much better with words and pictures than with words alone. My colleagues and I found this pattern in nine out of nine studies, yielding a median effect size of 1.50 (Mayer, 2009). I refer to this finding as the *multimedia principle*, and it is examined in detail by Butcher in Chapter 7.

The multimedia principle epitomizes the rationale for studying multimedia learning. There is reason to believe that, under certain circumstances, people learn more deeply from words and pictures than from words alone. For hundreds of years, the major format for instruction has been words, including lectures and books. In general, verbal modes of presentation have dominated the way we convey ideas to one another, and verbal learning has dominated education. Similarly, verbal learning has been the major focus of educational research.

With the recent advent of powerful computer graphics and visualization technologies, instructors have the ability to supplement verbal modes of instruction with pictorial modes of instruction. Advances in computer technology have led to an explosion in the availability of visual ways of presenting material, including large libraries of static images as well as compelling dynamic images in the form of animations and video. In light of the power of computer graphics, it may be useful to ask whether it is time to expand instructional messages beyond the purely verbal. What are the consequences of adding pictures to words? What happens when instructional messages involve both verbal and visual modes of learning? What affects the way that people learn from words and pictures? In short, how can multimedia presentations foster meaningful learning? These are the kinds of questions addressed in this handbook.

The case for multimedia learning is based on the idea that instructional messages should be designed in light of how the human mind works. Let's assume that humans have two information processing systems – one for verbal material and one for visual material, as described more fully in Part I. Let's also acknowledge that the major format for presenting instructional material is verbal. The rationale for multimedia presentation – that is, presenting material in words and pictures – is that it takes advantage of the full capacity of humans for processing information. When we present material only in the verbal mode, we are ignoring the potential contribution of our capacity to also process material in the visual mode.

Why might two channels be better than one? Two possible explanations are the quantitative rationale and the qualitative rationale. The quantitative rationale is that more material can be presented on two channels than on one channel – just as more traffic can travel on two lanes than on one lane. In the case of explaining how a bicycle tire pump works, for example, the steps in the process can be presented in words or can be depicted in illustrations.

Presenting both is like presenting the material twice – giving the learner twice as much exposure to the explanation. While the quantitative rationale makes sense as far as it goes, I reject it mainly because it is incomplete. In particular, I take exception to the assumption that the verbal and visual channels are equivalent – that is, that words and pictures are simply two equivalent ways of presenting the same material.

In contrast, the qualitative rationale is that words and pictures, while qualitatively different, can complement one another and that human understanding is enhanced when learners are able to mentally integrate visual and verbal representations. As you can see, the qualitative rationale assumes that the two channels are not equivalent; words are useful for presenting certain kinds of material – perhaps representations that are more abstract and require more effort to translate – whereas pictures are more useful for presenting other kinds of material – perhaps more intuitive, more natural representations. In short, one picture is not necessarily the same as a thousand words (or any number of words).

The most intriguing aspect of the qualitative rationale is that understanding occurs when learners are able to build meaningful connections between pictorial and verbal representations – such as being able to see how the words "the inlet valve opens" relate to the forward motion of the inlet valve in the cylinder of the pump. In the process of trying to build connections between words and pictures, learners are able to create a deeper understanding than from words or pictures alone. This idea is at the heart of the theories of multimedia learning described in Part I.

In summary, the rationale for the study of multimedia learning is that students may learn more deeply from words and pictures than from words alone. Thus, a primary purpose of this handbook is to explore the proposal that adding pictures to words may promote greater understanding than simply presenting words alone. However, not all pictures are equally effective. It is important to understand how best to incorporate pictures with words. Just because technologies are available that allow for state-of-the-art visualizations, this does not mean that instructors are well advised to use them. What is needed is a research-based understanding of how people learn from words and pictures and how to design multimedia instruction that promotes learning.

What Is the Research Base for Multimedia Learning?

Although research on verbal learning has a long and fruitful history in psychology and education, corresponding research on multimedia learning is just beginning to flourish. This second edition of *The Cambridge Handbook of Multimedia Learning* remains the world's first and most comprehensive summary of research on multimedia learning. In an attempt to organize the research base in multimedia learning, it is divided into five parts.

Part I – "Theoretical Foundations" – contains chapters that describe theories that are relevant to multimedia learning and that have had the greatest impact on research: Sweller's cognitive load theory (Chapter 2), Mayer's cognitive theory of multimedia learning (Chapter 3), Schnotz's integrative model of text and picture comprehension (Chapter 4), and van Merriënboer's four-component instructional design model for multimedia learning (Chapter 5).

Part II – "Basic Principles of Multimedia Learning" – begins with a chapter documenting questionable principles of multimedia learning, that is, principles that are commonly accepted but for which supporting evidence is lacking (Chapter 6, by Clark and Feldon). The remaining chapters explore the research evidence concerning basic principles for designing multimedia learning environments:

Multimedia principle: People learn better from words and pictures than from words alone (Chapter 7, by Butcher).

Split-attention principle: People learn better when words and pictures are physically and temporally integrated (Chapter 8, by Ayres and Sweller), similar to Mayer's spatial contiguity and temporal contiguity principles (Chapter 12).

Modality principle: People learn better from graphics and narration than from graphics and printed text (Chapter 9, by Low and Sweller), similar to Mayer's modality principle (Chapter 13).

Redundancy principle: People learn better when the same information is not presented in more than one format (Chapter 10, by Kalyuga and Sweller), similar to Mayer's redundancy principle (Chapter 12).

Signaling principle: People learn better when cues are added that highlight the key information and its organization (Chapter 11, by van Gog), similar to Mayer's signaling principle (Chapter 12).

Coherence, signaling, spatial contiguity, temporal contiguity, and redundancy principles: People learn better when extraneous material is excluded rather than included, when cues are added that highlight the organization of the essential material, and when corresponding words and pictures are presented near rather than far from each other on the screen or page or in time, and people learn better from graphics and narration than from graphics, narration, and on-screen text (Chapter 12, by Mayer and Fiorella).

Segmenting, pre-training, and modality principles: People learn better when a multimedia message is presented in learner-paced segments rather than as a continuous unit, people learn better from a multimedia message when they know the names and characteristics of the main concepts, and people learn better from a multimedia message when the words are spoken rather than written (Chapter 13, by Mayer and Pilegard).

Personalization, voice, embodiment, and image principles: People learn better when the words of a multimedia presentation are in conversational style rather than formal style, when the words are spoken in a

standard-accented human voice rather than a machine voice or foreign-accented human voice, and when on-screen agents display humanlike gestures and movements; but people do not necessarily learn better when the speaker's image is on the screen (Chapter 14, by Mayer).

Part III – "Advanced Principles of Multimedia Learning" – contains chapters that explore the research evidence for advanced principles of multimedia learning:

Guided discovery principle: People learn better when guidance is incorporated into discovery-based multimedia environments (Chapter 15, by de Jong and Lazonder).
Worked examples principle: People learn better when they receive worked examples in initial skill learning (Chapter 16, by Renkl).
Self-explanation principle: People learn better when they are encouraged to generate self-explanations during learning (Chapter 17, by Wylie and Chi).
Drawing principle: People learn better when they create drawings as they read explanative text (Chapter 18, by Leutner and Schmeck).
Feedback principle: People learn better from multimedia lessons when they receive explanative feedback on their performance (Chapter 19, by Johnson and Priest).
Multiple representation principle: There are circumstances under which people learn better from multiple representations (Chapter 20, by Ainsworth).
Learner control principle: People do not necessarily learn better when they have more control of the selection and organization of the material (Chapter 21, by Scheiter).
Animation principle: People do not necessarily learn better from animation than from static diagrams (Chapter 22, by Lowe and Schnotz).
Collaboration principle: People can learn better with collaborative online learning activities (Chapter 23, by Kirschner, Kirschner, and Janssen).
Prior knowledge principle: Instructional design principles that enhance multimedia learning for novices may hinder multimedia learning for more expert learners (Chapter 24, by Kalyuga).
Working memory principle: The effectiveness of instructional design principles depends on the learner's working memory capacity (Chapter 25, by Wiley, Sanchez, and Jaeger).

Part IV – "Multimedia Learning of Cognitive Processes" – takes a somewhat different cut by examining research on how to design multimedia learning to support higher-level cognition. The chapters summarize research on multimedia learning of cognitive skills (Chapter 26, by Lajoie), metacognitive strategies (Chapter 27, by Azevedo), and reasoning about complex systems (Chapter 28, by Hegarty).

Finally, the chapters in Part V – "Multimedia Learning in Advanced Computer-Based Contexts" – examine multimedia learning research involving emerging technologies. The chapters summarize research on multimedia learning with advanced technologies that have generated the most research, such as intelligent tutoring systems (Chapter 29, by Nye, Graesser, and Hu), simulations and microworlds (Chapter 30, by Plass and Schwartz), games (Chapter 31, by Tobias et al.,), video (Chapter 32, by Derry, Sherin, and Sherin), multiple sources (Chapter 33, by Rouet and Britt), and e-courses (Chapter 34, by Clark).

In all of the chapters the focus is on empirical research evidence, including implications of research for theory and practice. Overall, each chapter is intended to showcase the research base in a sub-area of multimedia learning, note its limitations, and offer suggestions for future research.

What's New in the Second Edition?

Although the general goals remain the same (i.e., to take an evidence-based approach to the design of multimedia instruction), there are five major changes in this second edition of the *Handbook*: an increase in the research base, the addition of new topics, a broadening of contexts of studying multimedia learning, an identification of boundary conditions, and a refinement of theory.

Increase in the research base

The second edition reflects the strong growth of the empirical research base in the field of multimedia learning, with many new references beyond those found in the previous edition. The book contains all of the basic principles of multimedia learning (i.e., multimedia, split attention, modality, redundancy, signaling, coherence, spatial contiguity, temporal contiguity, segmenting, pre-training, modality, personalization, voice, and image) and most of the advanced principles of multimedia learning (i.e., guided discovery, worked examples, self-explanation, collaboration, and prior knowledge) found in the first edition, but the principles are now informed by a much richer evidence base.

In some basic multimedia principles, the research base has more than doubled since the publication of the first edition in 2005. For example, in Chapter 12 on the coherence, signaling, redundancy, spatial contiguity, and temporal contiguity principles, the total number of experimental comparisons across all five principles in the first edition was 40, compared with 99 in the second edition, reflecting an increase of more than 100%. Similarly, in Chapter 13 on the segmenting, pre-training, and modality principles, the total number of experimental comparisons across all three