CHAPTER 3

Cognitive Theory of Multimedia Learning

Richard E. Mayer
University of California, Santa Barbara

Abstract

A fundamental hypothesis underlying research on multimedia learning is that multimedia instructional messages that are designed in light of how the human mind works are more likely to lead to meaningful learning than those that are not. The cognitive theory of multimedia learning (CTML) is based on three cognitive science principles of learning: the human information processing system includes dual channels for visual/pictorial and auditory/verbal processing (i.e., dual-channels assumption); each channel has limited capacity for processing (i.e., limited capacity assumption); and active learning entails carrying out a coordinated set of cognitive processes during learning (i.e., active processing assumption). The cognitive theory of multimedia learning specifies five cognitive processes in multimedia learning: selecting relevant words from the presented text or narration, selecting relevant imago from the presented illustrations, organizing the selected words into a coherent verbal representation, organizing selected images into a coherent pictorial representation, and integrating the pictorial and representations and prior Multimedia instructional messages should be designed to prime these processes.

The Case for Multimedia Learning

What is the rationale for a theory of multimedia learning? People learn more deeply from words and pictures than from words alone. This assertion — which can be called the *multimedia principle* — underlies much of the interest in multimedia learning. For thousands of years, words have been the major format for instruction — including spoken words, and within the last few hundred years, printed words. Today, thanks to further technological advances, pictorial forms of instruction are becoming widely available, including dazzling computer-based graphics. However, simply adding pictures to words does not guarantee an improvement in learning — that is, all multimedia presentations are not equally effective. In this chapter I explore a theory aimed at understanding how to use words and pictures to improve human learning.

A fundamental hypothesis underlying research on multimedia learning is that multimedia instructional messages that are designed in light of how the human mind works are more likely to lead to meaningful learning than those that are not. For the past 15 years my colleagues and I at the University of California, Santa Barbara have been engaged in a sustained effort to construct an evidenced-based theory of multimedia learning that can guide the design of effective multimedia instructional messages (Mayer 2001, 2002, 2oo3a; Mayer & Moreno, 2003).

What is a multimedia instructional message? A multimedia instructional message is a communication containing words and pictures intended to foster learning. The communication can be delivered using any medium, including paper (i.e., book-based communications) or computers (i.e., computer-based communications). Words can include printed words (such as you are now reading) or spoken words (such as in a narration); pictures can include static graphics — such as illustrations or photos — or dynamic graphics — such as animation or video clips. This definition is broad enough to include textbook chapters, online lessons containing animation and narration, and interactive simulation games. For example, Figure 3.1 presents frames from a narrated animation on lightning formation, which we have studied in numerous experiments (Mayer, 2001).

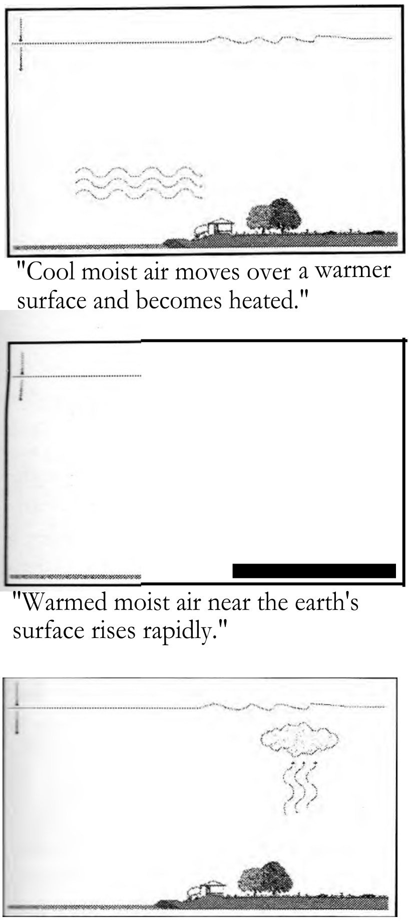
Learning can be measured by tests of retention (i.e., remembering the presented information) and transfer (i.e., being able to use the information to solve new problems). Our focus is on transfer because we are mainly interested in how words and pictures can be used to promote understanding. In short, transfer tests can help tell us how well people understand what they have learned. We are particularly interested in the cognitive processes by which people construct meaningful learning outcomes from words and pictures.

What is the role of a theory of learning in multimedia design? Much of the work presented in this handbook is based on the rpremise that the design of multimedia instructional messages should be compatible with how people learn. In short, the design of multimedia instructional messages should be sensitive to what we know about how people process information. The cognitive theory of multimedia learning represents an attempt to help accomplish this goal by describing how people learn from words and pictures, based on consistent empirical research evidence c.e (e.g., Mayer, 2001, 2002, 2oo3a; Mayer & Moreno, 2003) and on consensus principles in cognitive science (e.g., Bransford, Brown, & Cocking, 1999; Lambert & McCombs, 1998; Mayer, zoo3b).

In building the cognitive theory of multimedia learning my colleagues and I were guided by four criteria: *theoretical plausibility* – the theory is consistent with cognitive science principles of learning; *testability—the* theory yields predictions that can be tested in scientific research; *empirical plausibility* – the theory is consistent with empirical research evidence on multimedia learning; and – the theory is relevant to educational needs for improving the design of multimedia instructional messages. In this chapter, I describe the cognitive theory of multimedia learning, which is intended to meet these criteria. In particular, I summarize three underlying assumptions of the theory derived from cognitive science; describe ree memory stores, cognitive processes, and five o of representation in the theory; and then provide examples and a conclusion.

Three Assumptions of the Cognitive Theory of Multimedia Learning

Decisions about how to design a multimedia message always reflect an_underlyir u2Dception of how people learn — even when the underlying theory of learning is not stated. In short, the design of multimedia messages



"Cool moist air moves over a warmer surface and becomes heated."

"Warmed moist air near the earth's surface rises rapidly."

"As the air in this updraft cools, water vapor condenses into water droplets and forms a cloud."

Figure 3.1. Selected frames from a narrated animation on lightning formation.

is influenced by the designer's conception of how the human mind works.

, when a multimedia presentation consists of a screen overflowing with multicolored words and images — flashing and moving about — this reflects the designer's conception of human learning. The designer's conception is that human learners possess a single-channel, unlimited capacity, **and** passive-processing system. First, by not taking advantage of auditory modes of presentation, this design is based on a single-channel assumption — all information enters the cognitive system in the same way regardless of its modality. It follows that it does not matter which modality is used to present information — such as presenting words as sounds or text — just as long as the information is presented. Second, by presenting so much information, this design is based on an unlimited capacity assumption — humans can handle an unlimited amount of material. It follows that the designer's job is to present information to the learner. Third, by presenting many isolated pieces of information, this design is based on a passive processing assumption — humans act as tape recorders who add as much information to their memories as possible. It follows that learners do not need any guidance in organizing and making sense of the presented information.

What's wrong with this vision of learners as possessing a single-channel, unlimited capacity, and passive processing system? Current research in cognitive psychology paints a quite different view of how the human mind works (Bransford et al., 1999; Lambert 81 McCombs, 199 8; Mayer, 2 oo 3 b) . Thus, a difficulty with this commonsense conception of learning is that it conflicts with what is known about how people learn. In this section, I explore three assum tions underlying the cognitive theory o multimedia learning — *dual channels, limited capacity,* and *active cessing.* These assumptions are summarize in Table 3.1.

Dual-Channel Assumption

The dual-channel assumption is that humans possess separate information process-1 ing channels for visually represented material and auditorily represented material. The dual-channel assumption is incorporated into the cognitive theory of multimedia learning by proposing that the human information-processing system contains an auditory/verbal channel and a visual/pictorial channel. When information is presented to the eyes (such as illustrations, animations, video, or on-screen text), humans begin by processing that information

Table 3.1. Three Assumptions of a Cognitive Theory of Multimedia Learning

Assumption	*Description*	*Related citations*
Dual channels	Humans possess separate channels for processing visual and auditory information	Paivio (1986), Baddeley (1986, 1999)
Limited capacity	Humans are limited in the amount of information that can be processed in each channel at one time	Bacldeley (1986, 1999), Chandler & Sweller (1991)
Active processing	Humans engage in active learning by attending to relevant incoming information, organizing selected information into coherent mental representations, and integrating mental representations with other knowledge	Mayer (2001), Wittrock (198

in the v
ı ra
sounds), humans begin by processin that information in the auditory channel. The concept of separate information processing channels has a long history in cognitive psychology and currently is most closely associated with Paivio's dual-coding theory (Clark 8/ Paivio, 1991; Paivio, 1986) and Baddeley's model of working memory (Baddeley, 1986, 1999).

WHAT IS PROCESSED IN EACH CHANNEL?

There are two ways of conceptualizing the differences between the two channels - one based on ı
other based on *SefisWy modalities.* The
(presentation-mode -a
whether the presented stimulus is verbal \(such as spoken or printed words) or non-Verbal (such as pictures, video, animation, or background sounds). According to the presentation-mode approach, one channel processes verbal material and the other channel processes pictorial material and nonverbal sounds. This conceptualization is most consistent with Paivio's (1986) distinction between verbal and nonverbal systems.

j In contrast, the sensory-modality approach focuses on whether learners initially process the presented materials through their eyes (e.g., for pictures, video, animation, or printed words) or ears (e.g., for spoken words or background sounds). According to the sensory-modality approach, one channel processes visually represented material and the other channel processes auditorily represented material. This conceptualization is most consistent with Baddeley's (1986, 1999) distinction between the visuospatial sketchpad and the phonological (or articulatory) loop.

Whereas the presentation-mode approach focuses on the format of the stimulus-as-presentgd (i.e., verbal or nonverbal), the sensory-modality approach focuses on the stimulus-as-rearesented in working memory (i.e., auditory or visual). The major difference concerning multimedia learning rests in the processing of printed words (i.e., on-screen text) and background sounds. On-screen text is initially processed in the verbal channel in the presentation-mode approach but in the visual channel in the sensory-modality approach. Background sounds, including nonverbal music, are initially processed in the nonverbal channel in the presentation-mode approach but in the auditory channel in the sensory-mode approach.

For purposes of the cognitive theory of multimedia learning, I have opted for a compromise in which I use the sensory-modality approach to distinguish between visually presented material (e.g., pictures, animations, video, and on-screen text) and auditorily presented material (e.g., narration and background sounds) as well as a presentation-mode approach to distinguish between the construction of pictorially basedBaddeleyally based models in working memory. However, additional research is needed to clarify the nature of the differences between the two channels.

WHAT IS THE RELATION BETWEEN THE CHANNELS?

Although information enters the human information system through one channel, learners may also be able to convert the representation for processing in the other channel. When learners are able to devote ad- - equate cognitive resources to the task, it is possible for information originally presented to one channel to also be represented in the other channel. For example, on-screen text may initially be processed in the visual channel because it is presented to the eyes, but an experienced reader may be able to mentally convert images into sounds, which are processed through the auditory channel. Similarly, an illustration of an object or event such as a cloud rising above the freezing level may initially be processed in the visual channel, but the learner may also be able to mentally construct the corresponding verbal description in the auditory channel. Conversely, a narration describing some event such as "the cloud rises above the freezing level" may initially be processed in the auditory channel because it is presented to the ears, but the learn& may also form a corresponding mental image that is processed in the visual channel. Cross-channel representations of the same stimulus play an important role in Paivio's (1986) dual-coding theory.

Limited Capacity Assumption

The second assumption is that humans are limited in the amount of information that can be processed in each channel at one time. When an illustration or animation is presented, the learner is able to hold only a few images in working memory at any one time, reflecting portions of the presented material rather than an exact copy of the presented material. For example, if an illustration or animation of a tire pump is presented, the learner may be able to focus on building mental images of the handle going down, the inlet valve opening, and air moving into the cylinder. When a narration is presented, the learner is able to hold only a few words in working memory at any one time, reflecting portions of the presented text rather than a verbatim recording. For example, if the spoken text is "When the handle is pushed down, the piston moves down, the inlet valve opens, the outlet valve closes, and air enters the bottom of cylinder," the learner may be able to hold the following verbal representations in auditory working memory: "handle goes up," "inlet valve opens," and "air enters cylinder." The conception of limited capacity in consciousness has a long history in psychology, and some modern examples are Baddeley's (1986, 1999) theory of working memory and Chandler and Sweller's (1991; Sweller, 1999) cognitive load theory.

WHAT ARE THE LIMITS ON COGNITIVE CAPACITY?

If we assume that each channel has limited processing capacity, it is important to know just how much information can be processed in each channel. The classic way to measure someone's cognitive capacity is to give a memory span test (Miller, 1956; Simon, 1980). For example, in a digit span test, I can read a list of digits at the rate of one digit per second (e.g., 8-7-5-3-9-6-4) and ask you to repeat them back in order. The longest list that you can recite without making an error is your memory span for digits (or digit span). Alternatively, I can show you a series of line drawings of simple objects at the rate of one per second (e.g., moon-pencil-comb-apple-chair-book-pig) and ask you to repeat them back in order. Again, the longest list you can recite without making an error is your memory span for pictures. Although there are individual differences, on average memory span is fairly small — approximately five to seven chunks.

With practice, of course, people can learn techniques for chunking the elements in the list, such as grouping the seven digits 8-7-5-3-9-6-4 into three chunks 875-39-64 (e.g., "eight seven five" pause "three nine" pause "six four"). In this way, the cognitive capacity remains the same (e.g., five to seven chunks) but more elements can be remembered within each chunk. Researchers have developed more refined measures of verbal and visual working memory capacity, but continue to show that human processing capacity is severely limited (Miyake & Shah, 1999).

HOW ARE LIMITED COGNITIVE RESOURCES ALLOCATED?

The constraints on our processing capacity force us to make decisions about which pieces of incoming information to pay attention to, the degree to which we should build connections among the selected pieces of information, and the degree to which we should build connections between selected pieces of information and our existing knowledge.m*etacog*nitivetech—niques for allocating, monitoring coordinating, and adjusting these limited cognitive resources. These strategies are at the heart of what Baddeley (1986, 1999) calls the *central* *cutive* — the system that controls the allocation of cognitive resources — and play a central role in modern theories of intelligence (Sternberg, 1990).

Active Processing Assumption

The third assumption is that humans actively engage in cognitive processing in order to construct a coherent mental repre- l sentation of their experiences. These active cognitive processes include paying attention, (organizing incoming information, and integrating incoming information with other . In short, humans are active processors who seek to make sense of multimedia presentations. This view of humans as active processors conflicts with a common view of humans as passive processors who seek to add as much information as possible to memory that is, as tape recorders who file copies of their experiences in memory to be retrieved later.

WHAT ARE THE MAJOR WAYS THAT KNOWLEDGE CAN BE STRUCTURED?

Active learning occurs when a learner applies cognitive processes to incoming material — processes that are intended to help the learner make sense of the material. The outcome of active cognitive processing is the construction of a coherent mental representation, so active learning can be viewed as a process of model building. A *mental model (or knowledge structure)* represents the key parts of the presented material and their relations. For example, in a multimedia presentation of how lightning storms develop, the learner may attempt to build a cause-and-effect system in which a change in one part of the system causes a change in another part. In a lesson comparing and contrasting two theories, construction of a mental model involves building a sort of matrix structure that compares the two theories along several dimensions.

If the outcome of active learning is the construction of a coherent mental representation, it is useful to explore some of the typical ways that knowledge can be structured. Some basic knowledge structures include *process, comparison, generalization, enumeration,* and *classification* (Chambliss & Calfee, 1998; Cook & Mayer, 1988). Process structures can be represented as cause-and-effect chains and consist of explanations of how some system works. An example is an explanation of how the human ear works. Comparison structures can be represented as matrices and consist of comparisons among two or more elements along several dimensions. An example is a comparison between how two competing theories of learning view the role of the learner, the role of the teacher, and useful types of instructional methods. Generalization structures can be represented as a branching tree and consist of a main idea with subordinate supporting details. An example is a chapter outline for a chapter explaining the major causes for the American Civil War. Enumeration structures can be represented as lists and consist of a collection of items. An example is the names of principles of multimedia learning listed in this handbook. Classification structures can be represented as hierarchies and consist of sets and subsets. An example is a biological classification system for sea animals.

Understanding a multimedia message often involves constructing one of these kinds

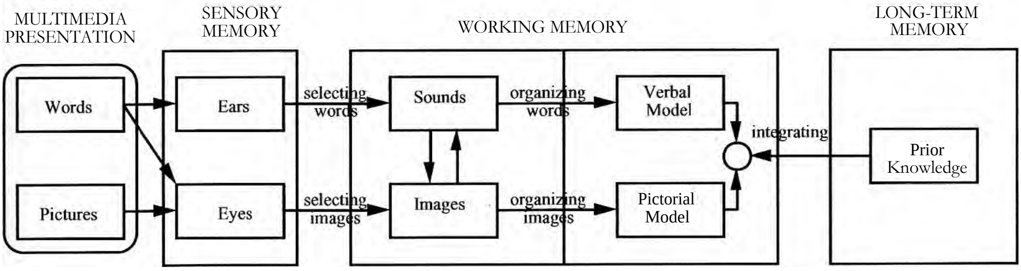


Figure 3.2. Cognitive theory of multimedia learning.

of knowledge structures. This assumption suggests two important implications for multimedia design: (1) the presented material should have a coherent structure and (2) the message should provide guidance to the learner for how to build the structure. If the material lacks a coherent structure — such as being a collection of isolated facts — the learner's model-building efforts will be fruitless. If the message lacks guidance for how to structure the presented material, the learner's model-building efforts may be overwhelmed. Multimedia design can be conceptualized as an attempt to assist learners in their model-building efforts.

WHAT ARE THE COGNITIVE PROCESSES INVOLVED IN ACTIVE LEARNING?

Three processes that are essential for active learning are selecting relevant material, organizing selected material, and integrating selected material with existing knowledge (Mayer, 1996, 2001; Wittrock, 1989). Selecting relevant material occurs when a learner pays attention to appropriate words and images in the presented material. This process involves bringing material from the outside into the working memory component of the cognitive system. Organizing selected material involves building structural relations among the elements — such as one of the five kinds of structures described in the preceding text. This process takes place within the working memory component of the cognitive system. Integrating selected material with existing knowledge involves building connections between incoming material and relevant portions of prior knowledge. This process involves activating knowledge in long-term memory and bringing it into working memory. For example, in a multimedia message on the cause of lightning, learners must pay attention to certain words and images, arrange them into a cause-and-effect chain, and relate the steps to prior knowledge such as the principle that hot air rises.

In sum, the implicit theory of learning underlying some multimedia messages is that learning is a single-channel, unlimited-capacity, passive-processing activity. In contrast, I offer a cognitive theory of multimedia learning that is based on three basic assumptions about how the human mind works — namely, that the human mind is a dual-channel, limited-capacity active-processing system.

Three Memory Stores in Multimedia Learning

Figure 3.2 presents a cognitive model of multimedia learning intended to represent the human information-processing system. The boxes represent memory stores, including sensory memory, working memory, and long-term memory. Pictures and words come in from the outside world as a multimedia presentation (indicated at the left side of the figure) and enter sensory memory through the eyes and ears (indicated in the sensory memory box). Sensory memory allows for pictures and printed text to be held as exact visual images for a very brief time period in a visual sensory memory (at the top) and for spoken words and other sounds to be held as exact auditory images for a very brief time period in an auditory sensory memory (at the bottom). The arrow from pictures to eyes corresponds to a picture being registered in the eyes, the arrow from words to ears corresponds to spoken text being registered in the ears, and the arrow from words to eyes corresponds to printed text being registered in the eyes.

The central work of multimedia learning takes place in working memory so let's focus there. Working memory is used for temporally holding and manipulating knowledge in active consciousness. For example, in reading this sentence you may be able to actively concentrate on only some of the words at one time, or in looking at Figure 3.2 you may be able to hold the images of only some of the boxes and arrows in your mind at one time. This kind of processing – that is, processing that involves conscious awareness — takes place in working memory. The left side of working memory represents the raw material that comes into working memory — visual images of pictures and sound images of words – so it is based on the two sensory modalities that I call visual and auditory. In contrast, the right side of working memory represents the knowledge constructed in working memory – pictorial and verbal models and links between them – so it is based on the two representation modes that I call pictorial and verbal. I use the term *pictorial model* to include spatial representations. The arrow from sounds to images represents the mental conversion of a sound (such as the spoken word *cat)* into a visual image (such as an image of a cat) – that is, when you hear the word "cat" you might also form a mental image of a cat. The arrow from images to sounds represents the mental conversion of a visual image (e.g., a mental picture of a cat) into a sound (e.g., the sound of the word "cat") – that is, you mentally hear the word *cat* when you see a picture of one. The major cognitive processing required for multimedia learning is represented by the arrows labeled *selecting images, selecting words, organizing images, organizing words,* and *integrating,* which are described in the next section.

Finally, the box on the right is labeled *long-term memory* and corresponds to the learner's storehouse of knowledge. Unlike working memory, long-term memory can hold large amounts of knowledge over long periods of time, but to actively think about material in long-term memory it must be brought into working memory (as indicated by the arrow from long-term memory to working memory).

Five Processes in the Cognitive Theory of Multimedia Learning

For meaningful learning to occur in a multimedia environment, the learner must engage in five cognitive processes: **(i)** selecting relevant words for processing in verbal working memory, (2.) selecting relevant images for processing in visual working memory, (3) organizing selected words into a verbal model, (4) organizing selected images into a pictorial model, and (5) integrating the verbal and pictorial representations with each other and with prior knowledge. Although I present these processes as a list, they do not necessarily occur in linear order, so a learner might move from process to process in many different ways. Successful multimedia learning requires that the learner coordinate and monitor these five processes.

Sekairiaelevant Words

The first labeled step listed in Figure 3.2 involves a change in knowledge representation from the external presentation of spoken words (e.g., computer-generated narration) to a sensory representation of sounds to an internal working memory representation of word sounds (e.g., some of the words in the narration). The input for this step is a spoken verbal message – that is, the spoken words in the presented portion of the multimedia message. The output for this step is a word sound base (called *sounds* in Figure 3.2) — that is, a mental representation in the learner's verbal working memory of selected words or phrases.

The cognitive process mediating this change is called *selecting relevant words* and involves paying attention to some of the words that are presented in the multimedia message as they pass through auditory sensory memory. If the words are presented as speech, this process begins in the auditory channel (as indicated by the arrows from *words* to *ears* to *sounds).* However, if the words are presented as on-screen text or printed text, this process begins in the visual channel (as indicated by the arrow from *words to eyes)* and later may move to the auditory channel if the learner mentally articulates the printed words (as indicated by the arrow from *images* to *sounds* in the left portion of working memory). The need for selecting only part of the presented message occurs because of capacity limitations in each channel of the cognitive system. If the capacity were unlimited, there would be no need to focus attention on only part of the verbal message. Finally, the selection of words is not arbitrary. The learner must determine which words are most relevant – an activity that is consistent with the view of the learner as an active sense maker.

For example, in the lightning lesson, one segment of the multimedia presentation contains the words, "Cool moist air moves over a warmer surface and becomes heated," the next segment contains the words, "Warmed moist air near the earth's surface rises rapidly," and the next segment has the words, 'As the air in this updraft cools, water vapor condenses into water droplets and forms a cloud." When a learner engages in the selection process, the result may be that some of the words are represented in verbal working memory – such as, "Cool air becomes heated, rises, forms a cloud."

Selecting Relevant Images

The second step involves a change in knowledge representation from the external presentation of pictures (e.g., an animation segment or an illustration) to a sensory representation of unanalyzed visual images to an internal representation in working memory (e.g., a visual image of part of the animation or illustration). The input for this step is a pictorial portion of a multimedia message that is held briefly in visual sensory memory. The output for this step is a visual image base (called *images* in Figure 3.2) — a mental representation in the learner's working memory of selected images.

The cognitive process underlying this change – *selecting relevant images* – involves paying attention to part of the animation or illustrations presented in the multimedia message. This process begins in the visual channel, but it is possible to convert part of it to the auditory channel (e.g., by mentally narrating an ongoing animation). The need to select only part of the presented pictorial material arises from the limited processing capacity of the cognitive system. It is not possible to process all parts of a complex illustration or animation so learners must focus on only part of the incoming pictorial material. Finally, the selection process for images – like the selection process for words – is not arbitrary because the learner must judge which images are most relevant for making sense out of the multimedia presentation.

In the lightning lesson, for example, one segment of the animation shows blue-colored arrows – representing cool air – moving over a heated land surface that contains a house and trees; another segment shows the arrows turning red and traveling upward above a tree; and a third segment shows the arrows changing into a cloud with lots of dots inside. In selecting relevant images, the learner may compress all this into images of a blue arrow pointing rightward, a red arrow pointing upward, and a cloud. Details such as the house and tree on the surface, the wavy form of the arrows, and the dots in the cloud are lost.

Organizing *Words*

Once the learner has formed a word sound base from the incoming words of a segment of the multimedia message, the next step is to organize the words into a coherent representation – a knowledge structure that I call *a verbal model.* The input for this step is the word sound base – the word sounds selected from the incoming verbal message. The output for this step is a verbal model – a coherent (or structured) representation in the learner's working memory of the selected words or phrases.

The cognitive process involved in this change is *organizing selected words* in which the learner builds connections among pieces of verbal knowledge. This process is most likely to occur in the auditory channel and is subject to the same capacity limitations that affect the selection process. Learners do not have unlimited capacity to build all possible connections so they must focus on building a simple structure. The organizing process is not arbitrary, but rather reflects an effort at sense making – such as the construction of a cause-and-effect chain.

For example, in the lightning lesson, the learner may build causal connections between the selected verbal components: "First: cool air is heated; second: it rises; third: it forms a cloud." In mentally building a causal chain, the learner is organizing the selected words.

Organizing Selected *Images*

The process for organizing images parallels that for selecting words. Once the learner has formed an image base from the incoming pictures of a segment of the multimedia message, the next step is to organize the images into a coherent representation – a knowledge structure that I call a *pictorial model.* The input for this step is the visual image base – the images selected from the incoming pictorial message. The output for this step is a pictorial model – a coherent (or structured) representation in the learner's working memory of the selected images.

This change from images to pictorial model requires the application of a cognitive process that I call *organizing selected images.* In this process, the learner builds connections among pieces of pictorial knowledge. This process occurs in the visual channel, which is subject to the same capacity limitations that affect the selection process. Learners lack the capacity to build all possible connections among images in their working memory, but rather must focus on building a simple set of connections. As in the process of organizing words, the process of organizing images is not arbitrary. Rather, it reflects an effort to build a simple structure that makes sense to the learner – such ass% cause-and-effect chain.

For example, in the lightning lesson, the learner may build causal connections between the selected images: The rightward-moving blue arrow turns into a rising red arrow, which turns into a cloud. In short, the learner builds causal links in which the first event leads to the second and so on.

ntegrating Word-Based and ĭmage-Based Representations

Perhaps the most crucial step in multimedia learning involves making connections between word-based and image-based representations. This step involves a change from having two separate representations – a pictorial model and a verbal model – to having an integrated representation in which corresponding elements and relations from one model are mapped onto the other. The input for this step is the pictorial model and the verbal model that the learner has constructed so far, and the output is an integrated model, which is based on connecting the two representations. In addition, the integrated model includes connections with prior knowledge.

I refer to this cognitive process as *integrating words and images* because it involves building connections between corresponding portions of the pictorial and verbal models as well as knowledge from long-term memory. This process occurs in visual and verbal working memory, and involves the coordination between them. This is an extremely demanding process that requires the efficient use of cognitive capacity. The process reflects the epitome of sense making because the learner must focus on the underlying structure of the visual and verbal representations. The learner can use prior knowledge to help coordinate the integration process, as indicated by the arrow from long-term memory to working memory.

Table 3.2. Five Cognitive Processes in the Cognitive Theory of Multimedia Learning

Process	*Description*
Selecting words	Learner pays attention to relevant words in a multimedia message to create sounds in working memory
Selecting images	Learner pays attention to relevant pictures in a multimedia message to creáleitliges in working memory
Organizing words	Learner builds connections among selected words to create a coherent verbal model in working memory
Organizing images	Learner builds connections among selected images to create a coherent pictorial model in working memory
Integrating	Learner builds connections between verbal and pictorial models and with prior knowledge

For example, in the lightning lesson, the learner must see the connection between the verbal chain – "First, cool air is heated; second, it rises; third, it forms a cloud" — and the pictorial chain – the blue arrow followed by the red arrow followed by the cloud shape. In addition, prior knowledge can be applied to the transition from the first to the second event by remembering that hot air rises.

The five cognitive processes in multimedia learning are summarized in Table 3.2. Each of the five processes in multimedia learning is likely to occur many times throughout a multimedia presentation. The processes are applied segment by segment rather than to the entire message as a whole. For example, in processing the lightning lesson, learners do not first select all relevant words and images from the entire passage, then organize them into verbal and pictorial models of the entire passage, and then connect the completed models with one another at the very end. Rather, learners carry out this procedure on small segments: they select relevant words and images from the first sentence of the narration and the first few seconds of the animation; they organize and integrate them; and then this set of processes is repeated for the next segment, and so on.

Five Forms of Representation

As you can see in Figure 3.2, there are five forms of representation for words and pictures, reflecting their stage of processing. To the far left, we begin with *words and pictures in the multimedia presentation,* that is, the stimuli that are presented to the learner. In the case of the lightning message shown in Figure 3.1, the words are the spoken words presented through the computer's speakers and the pictures are the frames of the animation presented on the computer's screen. Second, as the presented words and pictures impinge on the learner's ears and eyes, the next form of representation is *acoustic representations (or sounds) and iconic representations (or images) in sensory memory.* The sensory representations fade rapidly, unless the learner pays attention to them. Third, when the learner selects some of the words and images for further processing in working memory, the next form of representation is *sounds and images in working memory.* These are the building blocks for knowledge construction – including key phrases such as, "warmed air rises," and key images such as red arrows moving upward. The fourth form of representation results from the learner's construction of a *verbal model and pictorial model in working memory.* Here the learner has organized the material into coherent verbal and pictorial representations, and also has mentally integracreate images inally, the fifth form of representation is *knowledge in long-term memory,* which the learner uses for guiding the process of knowledge construction in working memory. Sweller (1999, and chapter 2, this volume) refers to this knowledge as *schemas.* After new knowledge is constructed in working memory, it is stored

Table 3.3. Five Forms of Representation in the Cognitive Theory of Multimedia Learning

Type of knowledge	*Location*	*Example*
Words and pictures	Multimedia presentation	Sound waves from computer speaker: "Warmed moist air.....
Acoustic and iconic representations	Sensory memory	Received sounds *in learner's* ears: "Warmed moist air......
Sounds and images	Working memory	Selected sounds: "warmed air rises"
Verbal and pictorial models	Working memory	Mental model of cloud formation
Prior knowledge	Long-term memory	Schema for differences in air pressure

in long-term memory as prior knowledge to be used in supporting new learning. The five forms of representation are summarized in Table 3.3.

Examples of How Three Kinds of Presented Materials Are Processed

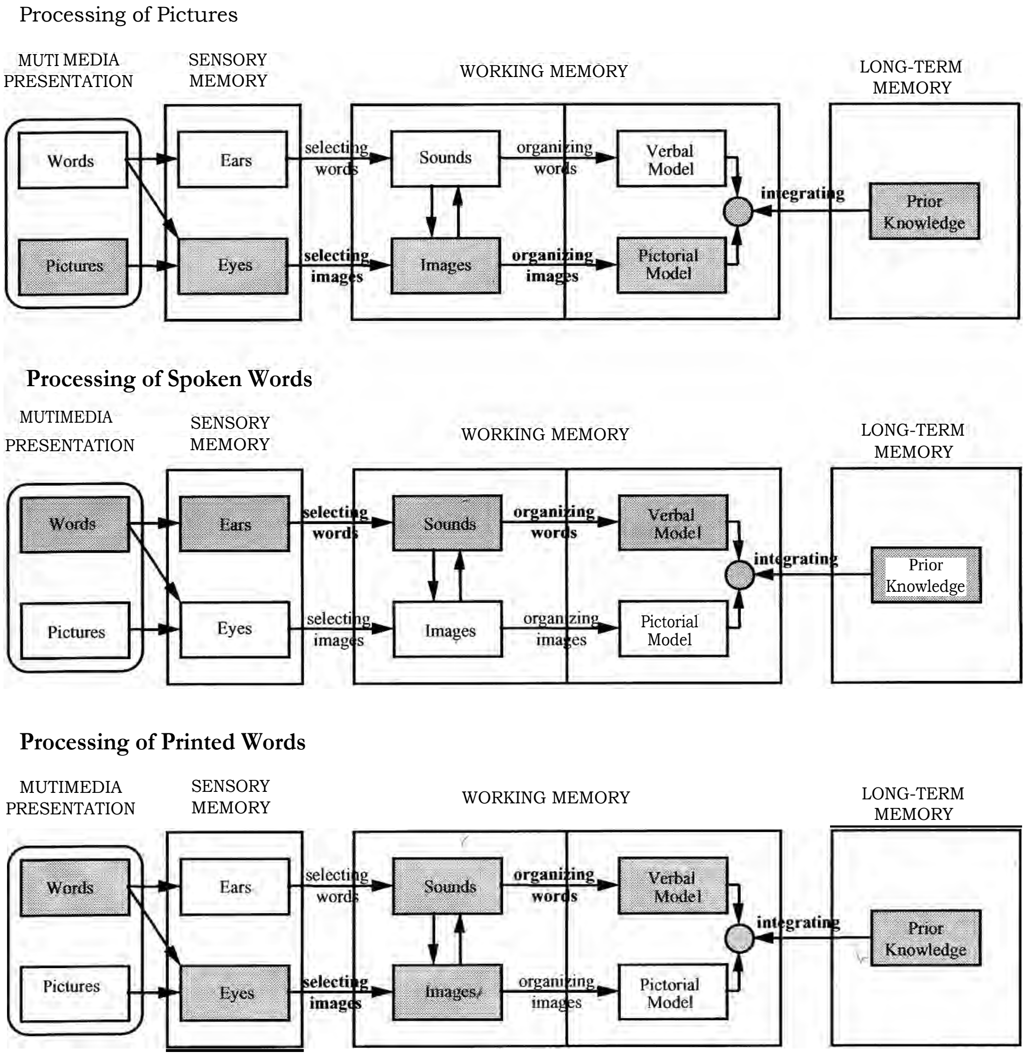
Let's take a closer look at how three kinds of presented materials are processed from start to finish according to the model of multimedia learning summarized in Figure 3.2: pictures, spoken words, and printed words. For example, suppose that a student clicks on an entry for lightning in a multimedia encyclopedia and is presented with a static picture of a lightning storm with a paragraph of on-screen text about the number of injuries and deaths caused by lightning each year. Similarly, suppose the student then clicks on the entry for lightning in another multimedia encyclopedia and is presented with a short animation along with narration describing the steps in lightning formation. In these examples, the first presentation contains static pictures and printed words whereas the second presentation contains dynamic pictures and spoken words.

Processing of Pictures

The top frame in Figure 3.3 shows the path for processing of pictures – indicated by thick arrows and darkened boxes. The first event – represented by the "pictures" box under "multimedia presentation" at the left side of Figure 3.3 — is the presentation of the lightning photograph from the first encyclopedia (i.e., a static picture) or the lighting animation from the second encyclopedia (i.e., a dynamic picture). The second event – represented by the "eyes" box under "sensory memory" – is that the pictures impinge on the eyes, resulting in a brief sensory image – that is for a brief time the student's eye beholds the photograph or the animation frames.

These first two events happen without much effort on the part of the learner, but next, the active cognitive processing begins – the processing over which the learner has some conscious control. If the student pays attention to the fleeting images coming from the eyes, parts of the images will become represented in working memory. This attentional processing corresponds to the arrow labeled "selecting images" and the resulting mental representation is labeled "images" under "working memory." Once working memory is full of image pieces, the next active cognitive processing involves organizing those pieces into a coherent structure – a process indicated by the "organizing images" arrow. The resulting knowledge representation is a *pictorial model,* that is, the student builds an organized visual representation of the main parts of a lightning bolt (from the first encyclopedia) or an organized set of images representing the cause-and-effect steps in lightning formation (from the second encyclopedia).

Finally, active cognitive processing is required to connect the new representation with other knowledge – a process indicated by the "integrating" arrow. For example, the


Processing of Pictures

Processing of Spoken Words

Processing of Printed Words

Figure 3.3. Processing pictures, spoken words, and printed words.

student may use prior knowledge about electricity to help include moving positive and negative charges in the mental representation of the lightning bolt or may use prior knowledge of electricity to help explain why the negative and positive charges are attracted to one another. In addition, if the learners have also produced a verbal model, they may try to connect it to the pictorial model – such as looking for how a phrase in the text corresponds to a part of the image. This processing results in an integrated learning outcome indicated by the circle under "working memory"

Processing of Spoken Words

The middle frame in Figure 3.3 shows the path for processing of spoken words – indicated by thick arrows and darkened boxes. When the computer produces narration (as indicated by the "words" box under "multimedia presentation") the sounds are picked up by the student's ears (as indicated by the "ears" box under "sensory memory"). For example, when the computer says, "The negatively charged particles fall to the bottom of the cloud, and most of the positively charged particles rise to the top," these words are picked up by the student's ears and held temporarily in auditory sensory memory. Next, active cognitive processing can take place. If the student pays attention to the sounds coming into the ears (as indicated by the arrow labeled "selecting words"), some of the incoming sounds will be selected for inclusion in the word sound base (indicated by the "sounds" box under "working memory"). For example, the resulting collection of words in working memory might include "positive top, negative bottom." The words in the word base are disorganized fragments, so the next step – indicated by the "organizing words" arrow – is to build them into a coherent mental structure – indicated by the "verbal model" box. In this process, the words change from being represented based on sound to being represented based on word meaning. The result could be a cause-effect chain for the steps in lightning formation. Lastly, the student may use prior knowledge to help explain the transition from one step to another and may connect words with pictures – such as connecting "positive top, negative bottom" with an image of positive particles in the top of a cloud and negative charges in the bottom. This process is labeled "integrating" and the resulting integrated learning outcome is indicated by the circle under "working memory."

Processing of Printed Words

So far, cognitive processing of pictures takes place mainly in the bottom channel of Figure 3.2 — that is the visual/pictorial channel – whereas the cognitive processing of spoken words takes place mainly in the top channel – that is, the auditory/verbal channel. However, the arrow from "images" to the "sounds" in working memory indicates that the learner can mentally create sounds corresponding to the visual image – such as thinking the word *wind* upon seeing wavy arrows in the animation. Similarly, the arrow from "sounds" to "images" in working memory indicates that the learner can mentally create images corresponding to the words — such as visualizing a plus sign when the narration says "positively charged particle."

The presentation of printed text in multimedia messages creates an information-processing challenge for the dual-channel system portrayed in Figure 3.2. For example, consider the case of a student who must read text and view an illustration. The words are presented visually so they must initially be processed through the eyes – as indicated by the arrow from "words" to "eyes." Then, the student may attend to some of the incoming words (as indicated by the "selecting images" arrow) and bring them into working memory as part of the images. Then, by mentally pronouncing the images of the printed words the student can get the words into the auditory/verbal channel – as indicated by the arrow from the images to the sounds. Once the words are represented in the auditory/verbal channel they are processed like the spoken words, as described previously. This path is presented in the bottom frame of Figure 3.3. As you can see, when verbal material must enter through the visual channel, the words must take a complex route through the system, and must also compete for attention with the illustration that the student is also processing through the visual channel. The consequences of this problem are addressed in chapters 9 and ii on the modality principle.

Conclusion

Historical Overview

The cognitive theory of multimedia learning has evolved within the body of research papers produced by my colleagues and me at the University of California, Santa Barbara (UCSB) over the past 15 years. Although the name has changed over the years, the underlying elements of the theory – that is, dual channels, limited capacity, and active processing – have remained constant. orne names used early in the research proram – such as "model of meaningful learng" (Mayer, 1989) and "cognitive conditions for effective illustrations" (Mayer & Gallini, 1990) — emphasized the active processing element. Other names used later – such as model" (Mayer & Anderson, 1991, 1992) and "dual-processing model of multimedia learning" (Mayer & Moreno, 1998; Mayer, Moreno, Boire, & Vagge, 1999) — emphasized the dual-channels element. Yet other names — such as "generative theory" (Mayer, Steinhoff, Bower, & Mars, 1995) and "generative theory of multimedia learning" (Mayer, 1997; Plass, Chun, Mayer, & Leutner, 1998) — emphasized all three elements. The current name, "cognitive theory of multimedia learning," was used in Mayer, Bove, Bryman, Mars, and Tapangco (1996), Moreno and Mayer (2000), and Mayer, Heiser, and Lonn (2001), and was selected for use in major reviews (Mayer, 2001, 2002, 2003a; Mayer & Moreno, 2003).

An early predecessor to the flowchart representation shown in Figure 3.2 in this chapter was a dual-coding model shown in Mayer and Sims (1994, Figure 1) which contained the same two channels and three of the same five cognitive processes, but lacked two of the cognitive processes and sensory memory. Mayer, Steinhoff, Bower, and Mars (1995, Figure 1) and Mayer (1997, Figure 3) presented an intermediate version that is almost identical to the flowchart shown in Figure 3.2 except that it lacked long-term memory and sensory memory. Finally, the current version of the flowchart appeared in Mayer, Heiser, and Lonn (2001), and was reproduced in subsequent reviews (Mayer, 2001, Figure 2; Mayer, 2002, Figure 7; Mayer, 2003a, Figure 2; Mayer & Moreno, 2003, Figure 1). Thus, the model has developed by adding components — both cognitive processes and mental representations — and clarifying their role. The result is the cognitive theory of multimedia learning that is represented in the flowchart in Figure 3.2 of this chapter.

Comparison With Related Theories

As can be seen in Figure 3.2, the cognitive theory of multimedia learning involves (a) two channels (i.e., visual and verbal), (b) limited processing capacity, (c) three kinds of memory stores, and (d) five cognitive processes (selecting words, selecting images, organizing words, organizing images, and integrating), and (e) five kinds of representations (i.e., presented words and pictures; sounds and images in sensory memory- selected sounds and images in working memory; verbal and pictorial models in working memory; and knowledge in long-term memory). The theory incorporates elements from classic information-processing models, such as *two channels* from Paivio's (1986) dual-coding theory, *limited processing* capacity from Baddeley's (1986, 1999) model of working memory, and a flowchart representation of *memory stores* and *cognitive processes* from Atkinson and Shiffirin (1968).

Key components of the cognitive theory of multimedia learning are consistent with other multimedia instructional design theories such as Sweller's (1999, 2003, chapter 2) cognitive load theory, and Schnotz and Bannert's (2003; Schnotz, chapter 4) integrated model of text and picture comprehension.

First, consider Sweller's (1999, 2003, chapter 2) cognitive load theory. Like the cognitive theory of multimedia learning, Sweller's (1999) cognitive load theory acknowledges "separate channels for dealing with auditory and visual material" (p. 138) and emphasizes that "we can hold few elements in working memory" (p. 4). Like the cognitive theory of multimedia learning, the architecture of the human information processing allows for several kinds of representations: elements in the presented material correspond to words and pictures in the multimedia presentation, elements in working memory correspond to verbal and pictorial models in working memory, and schemas in long-term memory correspond to knowledge in long-term memory. Cognitive load theory elaborates on the implications of limited working memory capacity for instructional design, and focuses on ways in which instruction imposes cognitive load on learners. However, it does not focus on the kinds of information processes involved in multimedia learning.

Second, consider Schnotz and Bannert's integrated model of text and picture comprehension as summarized in Figure 3.2 of

Schnotz and Bannert (2003). Like the cognitive theory of multimedia learning, Schnotz and Bannert's model emphasizes two channels, but unlike the cognitive theory of multimedia learning it does not emphasize limited capacity. All five cognitive processes are represented although with some differences in conceptualization: subsemantic processing corresponds to selecting words, perception corresponds to selecting images, semantic processing corresponds to organizing words, thematic selection corresponds to organizing images, and model construction/inspection corresponds to integrating. Four of the five representations are included although, again, with some differences in conceptualization: text and picture/diagram corresponds to words and pictures in the multimedia presentation; text surface representation and visual image correspond to sounds and images in working memory; propositional representation and mental model correspond to verbal model and pictorial model; and conceptual organization corresponds to knowledge in long-term memory.

In summary, the cognitive theory of multimedia learning is compatible and somewhat similar to other multimedia design theories. Sweller's (1999, 2003, chapter 2) cognitive load theory offers further elaborations on the role of limited capacity in instructional design for multimedia learning, and Schnotz and Bannert's (2003, Schnotz, chapter 4) offers further elaborations on the nature of mental representations in multimedia learning.

Future Directions

Although we have made progress in creating a cognitive theory of multimedia learning, much remains to done, particularly (a) in fleshing out the details of the mechanisms underlying the five cognitive processes and the five forms of representation, (b) in integrating the various theories of multimedia learning, and (c) in building a credible research base. First, more work is needed to understand and measure the basic constructs in theories of multimedia learning, such as determining how to measure cognitive load during learning, determining the optimal size of a chunk of presented information, or determining the way that a mental model is represented in the learner's memory. Second, there is a need to find consensus among theorists, such as reconciliation among cognitive load theory (Sweller, chapter 2), and the cognitive theory of multimedia learning (this chapter), the integrative model of text and picture comprehension, (Schnoz, chapter 4), the four-component instructional design model (Merriënboer & Kester, chapter 5), and related theories. Third, we have a continuing need to generate testable predictions from theories of multimedia learning and to test these predictions in rigorous scientific experiments. The best way to insure the usefulness of theories of multimedia learning is to have coherent research literature on which to base them.

Summary

In summary, multimedia learning takes place within the learner's information system - a system that contains separate channels for visual and verbal processing, a system with erious limitations on the capacity of ea c , and a system that requires coordinated cognitive processing in each channel for active learning to occur. In particular, multimedia learning is a demanding process that requires selecting relevant words and images; organizing them into coherent verbal and pictorial representations; and integrating the verbal and pictorial representations with each other and with prior knowledge. In the process of multimedia learning, material is represented in five forms: as words and pictures in a multimedia acoustic and iconic representations in sensory memory; sounds and images in working memory; verbal and pictorial models in working memory; and knowledge in long-term memory. The theme of this chapter is that multimedia messages should be designed to facilitate multimedia learning processes. Multimedia messages that are designed in light of how the human mind works are more likely to lead to meaningful learning than those that are not. This proposition is tested empirically in the chapters of this handbook.

Glossary

Cognitive theory of multimedia learning: A theory of how people learn from words and pictures, based on the idea that people possess separate channels for processing verbal and visual material (dual-channels assumption), each channel can process only a small amount of material at a time (limited-capacity assumption), and meaningful learning involves engaging in appropriate cognitive processing during learning (active-processing assumption).

Long-term memory: A memory store that holds large amounts of knowledge over long periods of time.

Multimedia instructional message: A communication containing words and pictures intended to foster learning.

Multimedia principle: People learn more deeply from words and pictures than from words alone.

Sensory memory: A memory store that holds pictures and printed text impinging on the eyes as exact visual images for a very brief period and that holds spoken words and other sounds impinging on the ears as exact auditory images for a very brief period.

Working memory: A limited-capacity memory store for holding and manipulating sounds and images in active consciousness.

Note

This chapter is based on chapter 3, "A Cognitive Theory of Multimedia Learning," in *Multimedia Learning* (Mayer, 2001). I appreciate the helpful comments of Jeroen van Merrienboer, Wolfgang Schnotz, and John Sweller.

References

Atkinson, R. C., & Shiffrin, R. M. (1968). Human memory: A proposed system and its control processes. In K. W. Spence (Ed.), *The psychology of learning and motivation* (pp. 89-195). New York: Academic Press.

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

Baddeley, A. D. (1999). *Human memory.* Boston: Allyn & Bacon.

Bransford, J. D., Brown, A. L., & Cocking, R. R. (1999). *How people learn.* Washington, DC: National Academy Press.

Chambliss, M. J., & Calfee, R. C. (1998). *Textbooks for learning.* Oxford, England: Blackwell.

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

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

Cook, L. K., & Mayer, R. E. (1988). Teaching readers about the structure of scientific text. *Journal of Educational Psychology,* 8o, 448—456.

Lambert, N. M., & McCombs, B. L. (1998). *How students leant.* Washington, DC: American Psychological Association.

Mayer, R. E. (1989). Systematic thinking fostered by illustrations in scientific text. *Journal of Educational Psychology, 8x,* 240-246.

Mayer, R. E. (1996). Learning strategies for making sense out of expository text: The SOI model for guiding three cognitive processes in knowledge construction. *Educational Psychology Review, 8,* 357-371.

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

Mayer, R. E. (2001). *Multimedia learning.* New York: Cambridge University Press.

Mayer, R. E. (2002). Multimedia learning. In B. H. Ross (Ed.), *The psychology of learning and motivation: Volume* 41 (pp. 85-139). San Diego, CA: Academic Press.

Mayer, R. E. (2003a). The promise of multimedia learning: Using the same instructional design methods across different media. *Learning and Instruction,* 12,125-141.

Mayer, R. E. (2003b). *Learning and instruction.* Upper Saddle River, NJ: Merrill Prentice Hall.

Mayer, R. E., & Anderson, R. B. (1991). Animations need narrations: An experimental test of the 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., Bove, W, Bryman, A., Mars, R., & Tapangco, L. (1996). When less is more: Meaningful learning from visual and verbal summaries of science textbook lessons. *Journal of Educational Psychology,* 88,64-73.

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

Mayer, R. E., Heiser, J., & Lonn, S. (2001). Cognitive constraints on multimedia learning: When presenting more material results in less understanding. *Journal of Educational Psychology,* 93, 187-198.

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

Mayer, R. E., & Moreno, R. (2003). Nine ways to reduce cognitive load in multimedia learning. *Educational Psychologist,* 38,43-52.

Mayer, R. E., Moreno, R., Boire, M., & Vagge, S. (1999). Maximizing constructivist learning from multimedia communications by minimizing cognitive load. *Journal of Educational Psychology,* 91, 638-643 .

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

Mayer, R. E., Steinhoff K., Bower, G., & Mars, R. (i 995). A generative theory of textbook design: Using annotated illustrations to foster meaningful learning of science text. *Educational Technology Research Se Development,* 43, 31-43.

Miller, G. A. (1956). The magic number seven, plus or minus two: Some limits on our capacity for processing information. *Psychological Review, 63,* 81-97.

Miyake, A., & Shah, P. (Eds.). (1999). *Models of working memory.* New York: Cambridge University Press.

Moreno, R., & Mayer, R. E. (2000). A coherence effect in multimedia learning: The case for minimizing irrelevant sounds in the design of multimedia instructional messages. *Journal of Educational Psychology,* 92, 117-125.

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

Plass, J. L., Chun, D. M., Mayer, R. E., & Leaner, D. (1998). Supporting visual and verbal learning preferences in a second-language multimedia learning environment. *Journal of Educational Psychology,* 90, 25-36 .

Schnotz, W., & Bannert, M. (2003). Construction and interference in learning from multiple representation. *Learning and Instruction, 13*, 141-156.

Simon, H. A., (1974). How big is a chunk? *Science, 183*, 482-488.

Sternberg, R. J. (1990). *Metaphors of mind: Conceptions of the nature of intelligence.* New York: Cambridge University Press.

Sweller, J. (1999). *Instructional design in technical areas.* Camberwell, Australia: ACER Press.

Sweller, J. (2003). Evolution of human cognitive architecture. In B. Ross (Ed.), *The psychology of learning and motivation* (Vol. 43, pp. 215-216). San Diego, CA: Academic Press.

Wittrock, M. C. (1989). Generative processes of comprehension. *Educational Psychologist, 24,* 345-376-