CHAPTER 1

Introduction to Multimedia Learning

Richard E. Mayer
University of California, Santa Barbara

Abstract

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 *The Cambridge Handbook of Multimedia Learning*[1]. Each of the 35 chapters in this handbook 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, and draw distinctions between two approaches to multimedia design, three metaphors 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 sitting in a room where images are presented on one or more screens and music or other sounds are presented using speakers – that is, multimedia as a "live" performance. Alternatively, you might think of sitting in front of 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 television screen while listening to the corresponding words, music, and sounds, or watching a PowerPoint

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

presentation along with listening to the speaker's corresponding commentary. Low-tech examples of multimedia include a "chalk-and-talk" presentation where 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.

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 using printed text or spoken text. By pictures, I mean that the material is presented in *pictorial form*, such as using static graphics, including illustrations, graphs, diagrams, maps, or photos, or using dynamic graphics, including animation or video. This definition is broad enough to include all of the scenarios I described in the previous 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.

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 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 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, 2001a). The presentation modes view is also the basis for Paivio's (1986) dual-code theory as well theories of multimedia learning presented in this handbook (Mayer, chapter 3; Schnotz, chapter 4; Sweller, chapter 2; van Merriënboer & Kester, chapter 5). 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 in this definition, *multimedia* refers to the presentation of words and pictures, whereas *learning* refers to the learner's construction of knowledge. The process by which people build mental representations from words and pictures is the focus of Mayer's cognitive theory of multimedia learning (Mayer, chapter 3), Sweller's cognitive load theory (Sweller, chapter 2), and Schnotz's integrative model of text and picture comprehension (Schnotz, chapter 4).

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 presentations in ways that help people build mental representations. The instructional design principles described in parts 2 and 3 of this handbook suggest ways of creating multimedia presentations intended to promote multimedia learning.

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, 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 in 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 and 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 transfer tests (Mayer & Anderson, 1991, 1992). In particular, when we focus on tests of problem-solving transfer – which are designed to measure the student's understanding of the presented material – students perform much better with words and pictures than from words alone. My colleagues and I found this pattern in nine out of nine studies, yielding a median effect size of 1.50 (Mayer, 2001a). I refer to this finding as the *multimedia principle* and it is examined in detail by Fletcher and Tobias 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 in 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 enabled 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

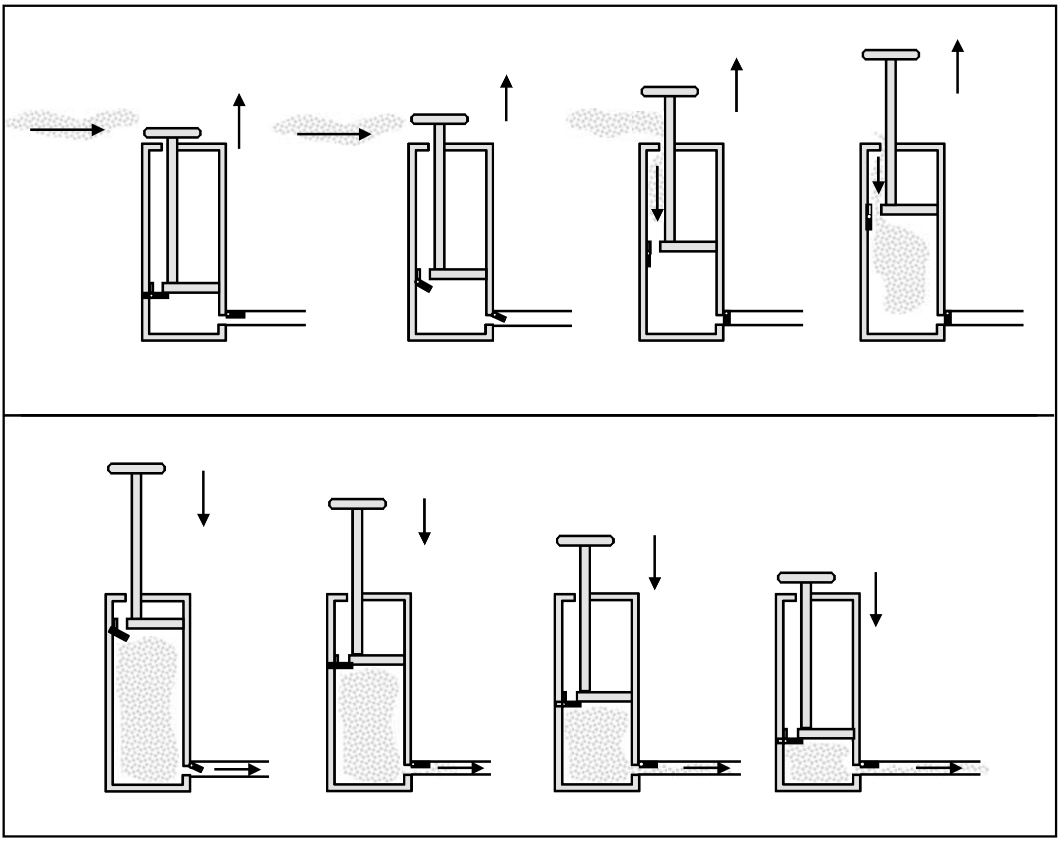

Figure 1.1. Frames from a pumps animation.

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 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 1 of this handbook. Let's also acknowledge that the major format for presenting instructional material is verbal. The rationale for multimedia presentations – 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 like more traffic can travel over two lanes than 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 am concerned about the assumption

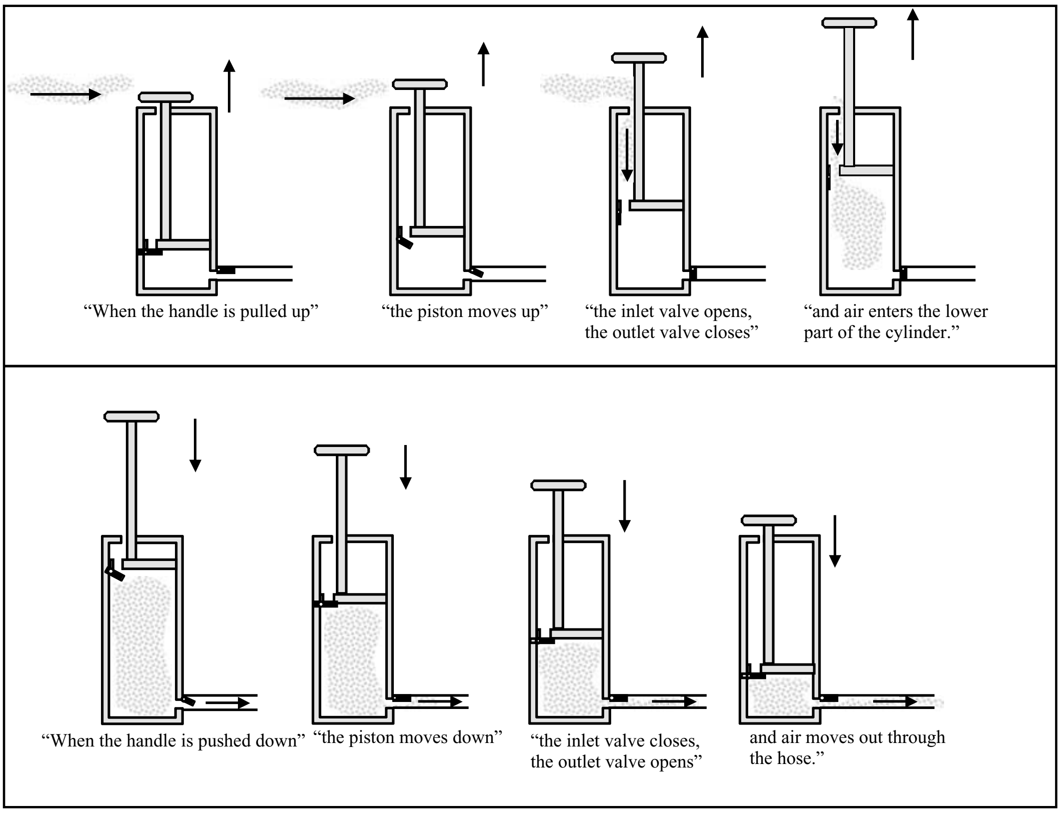

Figure 1.2. Frames from a pumps animation with corresponding narration.

that the verbal and visual channels are equivalent, that is, that words and pictures are simply two equivalent ways for 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 equivalent to 1,000 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 visual 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 1 of this handbook.

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, the motivation for this handbook is to explore the proposal that adding pictures to words may be a way of helping people understand better than by 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. The *Cambridge Handbook of Multimedia Learning* is the world's first comprehensive summary of research on multimedia learning. In an attempt to organize the research base in multimedia learning, this handbook is divided into five parts.

Part 1: Theoretical Foundations contains chapters that describe the theories of multimedia learning that have had the greatest impact on research: Sweller's cognitive load theory (CLT) (chapter 2), Mayer's cognitive theory of multimedia learning (CTML) (chapter 3), Schnotz's integrative model of text and picture comprehension (chapter 4), and van Merriënboer and Kester's four-component instructional design (4C-ID) model for multimedia learning (chapter 5).

Part 2: Basic Principles of Multimedia Learning begins with a chapter documenting mistaken principles of multimedia learning, that is, principles that are commonly accepted but for which supporting evidence is lacking (Clark & Feldon, chapter 6) The remaining chapters explore the research evidence concerning basic principles for how to design multimedia learning environments:

multimedia principle – People learn better from words and pictures than from words alone (Fletcher & Tobias, chapter 7).

split-attention principle – People learn better when words and pictures are physically and temporally integrated (Ayres & Sweller, chapter 8). This is similar to Mayer's (chapter 12) spatial contiguity and temporal contiguity principles.

modality principle – People learn better from graphics and narration than graphics and printed text (Low & Sweller, chapter 9). This is similar to Mayer's modality principle (chapter 11).

redundancy principle – People learn better when the same information is not presented in more than one format (Sweller, chapter 10). This is similar to Mayer's redundancy principle (chapter 12).

segmenting, pretraining, and modality principles – People learn better when a multimedia message is presented in learned-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 (Mayer, chapter 11).

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, 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 (Mayer, chapter 12).

personalization, voice, and image principles – People learn better when the words of a multimedia presentation are in conversational style rather than formal style and when the words are spoken in a standard-accented human voice rather than a machine voice or foreign-accented human voice; but people do not necessarily learn better when the speaker's image is on the screen (Mayer, chapter 13).

Part 3: 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 (de Jong, chapter 14).

worked-out example principle – People learn better when they receive worked-out examples in initial skill learning (Renkl, chapter 15).

collaboration principle – People can learn better with collaborative online learning activities (Jonassen, Lee, Yang, & Laffey, chapter 16).

self-explanation principle – People learn better when they are encouraged to generate self-explanations during learning (Roy & Chi, chapter 17).

animation and interactivity principles – People do not necessarily learn better from animation than from static diagrams (Betrancourt, chapter 18).

navigation principles – People learn better in hypertext environments when appropriate navigation aids are provided (Rouet & Pottelle, chapter 19).

site map principle – People can learn better in an online environment when the interface includes a map showing where the learner is in the lesson (Shapiro, chapter 20).

prior knowledge principle – Instructional design principles that enhance multimedia learning for novices may hinder multimedia learning for more expert learners (Kalyuga, chapter 21).

cognitive aging principle – Instructional design principles that effectively expand working memory capacity are especially helpful for older learners (Paas, van Gerven & Tabbers, chapter 22).

Part 4: Multimedia Learning in Content Areas takes a somewhat different cut by examining research on how to design multimedia learning environments in various content areas. The chapters summarize research on multimedia learning in content areas that have generated the most research on multimedia learning including reading (Reinking, chapter 23), history (Wiley & Ash, chapter 24), mathematics (Atkinson, chapter 25), chemistry (Kozma & Russell, chapter 26), meteorology (Lowe, chapter 27), complex physical systems (Hegarty, chapter 28), second language learning (Plass & Jones, chapter 29), and cognitive skills (Lajoie & Nakamura, chapter 30).

Finally, in *Part 5: Multimedia Learning in Advanced Computer-Based Contexts*, the chapters examine multimedia learning research involving emerging new technologies. The chapters summarize research on multimedia learning with advanced technologies that have generated the most research, such as animated pedagogical agents (Moreno, chapter 31); virtual reality (Cobb & Fraser, chapter 32); games, simulations, and microworlds (Rieber, chapter 33); hypermedia (Dillon & Jobst, chapter 34); and e-courses (Clark, chapter 35).

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

Two Approaches to Multimedia Learning: Technology Centered Versus Learner Centered

Multimedia represents a potentially powerful learning technology – that is, a system for enhancing human learning. A practical goal of research on multimedia is to devise design principles for multimedia presentations. In addressing this goal, it is useful to distinguish between two approaches to multimedia design – a technology-centered approach and a learner-centered approach.

Technology-Centered Approaches

The most straightforward approach to multimedia design is technology centered. Technology-centered approaches begin with the functional capabilities of multimedia and ask, "How can we use these capabilities in designing multimedia presentations?" The focus is generally on cutting-edge advances in multimedia technology, so technology-centered designers might focus on how to incorporate multimedia into emerging communications technologies such as wireless access to the World Wide Web or the construction of interactive multimedia representations in virtual reality. The kinds of research issues often involve media research (i.e., determining which technology is most effective in presenting information). For example, a media research issue is whether students learn as well from an online lecture – in which the student can see a lecturer in a window on the computer screen – as they can from a live lecture – in which the student is actually sitting in a classroom.

What's wrong with technology-centered approaches? A review of educational technologies of the twentieth century shows that the technology-centered approach generally fails to lead to lasting improvements in education (Cuban, 1986). For example, when the motion picture was invented in the early 20th century hopes were high that this visual technology would improve education. In 1922, the famous inventor Thomas Edison predicted that "the motion picture is destined to revolutionize our educational system and that in a few years it will supplant largely, if not entirely, the use of textbooks" (cited in Cuban, 1986, p. 9). Like current claims for the power of visual media, Edison proclaimed that "it is possible to teach every branch of human knowledge with the motion picture" (cited in Cuban, 1986, p. 11). In spite of the grand predictions, a review of educational technology reveals that "most teachers used films infrequently in their classrooms" (Cuban, 1986, p. 17). From our vantage point beyond the close the 20th century it is clear that the predicted educational revolution in which movies would replace books has failed to materialize.

Consider another disappointing example that may remind you of current claims for the educational potential of the World Wide Web. In 1932, Benjamin Darrow, founder of the Ohio School of the Air, proclaimed that radio could "bring the world to the classroom, to make universally available the services of the finest teachers, the inspiration of the greatest leaders . . . " (cited in Cuban, 1986, p. 19). His colleague, William Levenson, the director of the Ohio School of the Air predicted in 1945 that a "radio receiver will be as common in the classroom as the blackboard" and "radio instruction will be integrated into school life" (cited in Cuban, 1986, p. 19). As we rush to wire our schools and homes for access to the educational content of the Internet, it is humbling to recognize what happened to a similarly motivated movement for radio: "Radio has not been accepted as a full-fledged member of the educational community" (Cuban, 1986, p. 24).

Third, consider the sad history of educational television – a technology that combined the visual power of the motion picture with the worldwide coverage of radio. By the 1950s, educational television was touted as a way to create a "continental classroom" that would provide access to "richer education at less cost" (Cuban, 1986, p. 33). Yet, a review shows that teachers used television infrequently, if at all (Cuban, 1986).

Finally, consider the most widely acclaimed technological accomplishment of the 20th century – computers. The technology that supports computers is different from film, radio, and television, but the grand promises to revolutionize education are the same. Like current claims for the mind-enhancing power of computer technology, during the 1960s computer tutoring machines were predicted to eventually replace teachers. The first large-scale implementation occurred under the banner of computer-assisted instruction (CAI) in which computers presented short frames, solicited a response from the learner, and provided feedback to the learner. In spite of a large financial investment to support

CAI, sound evaluations showed that the two largest computer-based systems in the 1970s – PLATO and TICCIT – failed to produce better learning than traditional teacher-lead instruction (Cognition and Technology Group at Vanderbilt, 1996).

What can we learn from the humbling history of the 20th century's great educational technologies? Although different technologies underlie film, radio, television, and computer-assisted instruction, they all produced the same cycle. First, they began with grand promises about how the technology would revolutionize education. Second, there was an initial rush to implement the cutting-edge technology in schools. Third, from the perspective of a few decades later it became clear that the hopes and expectations were largely unmet.

What went wrong with these technologies that seemed poised to tap the potential of visual and worldwide learning? I attribute the disappointing results to the technology-centered approach taken by the promoters. Instead of adapting technology to fit the needs of human learners, humans were forced to adapt to the demands of cutting-edge technologies. The driving force behind the implementations was the power of the technology rather than an interest in promoting human cognition. The focus was on giving people access to the latest technology rather than helping people to learn through the aid of technology.

Are we about to replicate the cycle of high expectations, large-scale implementation, and disappointing results in the realm of multimedia technology? In my opinion, the answer to that question depends on whether or not we continue to take a technology-centered approach. When we ask, "What can we do with multimedia?" and when our goal is to "provide access to technology," we are taking a technology-centered approach with a 100-year history of failure.

Learner-Centered Approaches

Learner-centered approaches offer an important alternative to technology-centered approaches. Learner-centered approaches begin with an understanding of how the human mind works and ask, "How can we adapt multimedia to enhance human learning?" The focus is on using multimedia technology as an aid to human cognition. Research questions focus on the relation between design features and the human information processing system, such as, comparing multimedia designs that place light or heavy loads on the learner's visual information processing channel. The premise underlying the learner-centered approach is that multimedia designs that are consistent with the way the human mind works are more effective in fostering learning than those that are not. This premise is the central theme of part 1 of this handbook, which lays out theories of multimedia learning.

Norman (1993, p. xi) eloquently makes the case for a learner-centered approach to technology design, which he refers to as *human-centered technology*: "Today we serve technology. We need to reverse the machine-centered point of view and turn it into a person-centered point of view: Technology should serve us." Consistent with the learner-centered approach, Norman (1993, p. 3) shows how "technology can make us smart" – that is, technology can expand our cognitive capabilities. Norman (1993, p. 5) refers to tools that aid the mind as *cognitive artifacts*: "anything invented by humans for the purpose of improving thought or action counts as an artifact." Examples include mental tools such as language and arithmetic as well as physical tools such as paper and pencils. As the 20th century's most important new cognitive artifact, computer technology represents a landmark invention that has the potential to assist human cognition in ways that were previously not possible.

Norman's (1993, p. 9) assessment is that "much of science and technology takes a machine-centered view of the design of machines" so that "the technology that is intended to aid human cognition . . . more often interferes and confuses." In contrast, Norman's (1993, p. 12) vision of a learner-centered approach to technology design is that "technology . . . should complement human abilities, aid those activities for which

Table 1.2. Two Approaches to the Design of Multimedia Instruction

Design Approach	*Starting Point*	*Goal*	*Issues*
Technology centered	Capabilities of multimedia technology	Provide access to information	How can we use cutting-edge technology in designing multimedia instruction?
Learner centered	How the human mind works	Aid human cognition	How can we adapt multimedia technology to aid human cognition?

we are poorly suited, and enhance and help develop those for which we are ideally suited." The design of multimedia technology to promote human cognition represents one exemplary component in the larger task of creating what Norman (1993, p. xi) calls "things that make us smart."

In his review of computer technology, Landauer (1995, p. 3) proclaims that "the computer and information revolution is widely predicted to be as consequential as the industrial revolution of the previous two centuries." Further, he describes two major phases in the use of computer technology – *automation* and *augmentation*. In the automation phase, computers are used to replace humans on certain tasks ranging from robots in manufacturing to imaging devices (e.g., CAT scans and MRIs) in medicine to computer-based switching in telecommunications. However, Landauer (1995, p. 6) observes that the automation phase "is running out of steam" because almost all of the easy to automate tasks have been computerized.

The second phase of computer application – augmentation – involves the use of computers to enhance human performance on various cognitively complex tasks. Augmentation involves designing computer systems "to act as assistants, aids, and power tools" (Landauer, 1995, p. 7). However, Landauer (1995, p. 7) is disappointed with progress in the augmentation phase: "It is here . . . that we have failed." A major challenge in making the augmentation phase work involves the learner-centered design of computer-based technologies: "They are still too hard to use" (Landauer, 1995, p. 7). The design of multimedia learning environments that promote meaningful human learning is an example of using computers to augment or aid human cognition – and thus one element in Landauer's augmentation phase.

The differences between the technology-centered and learner-centered approaches to multimedia design are summarized in Table 1.2.

Three Metaphors of Multimedia Learning: Response Strengthening, Information Acquisition, and Knowledge Construction

In making decisions about how to design or select a multimedia learning environment, you may be influenced by your underlying conception of learning. Table 1.3 compares three views of multimedia learning – *multimedia learning as response strengthening, multimedia learning as information acquisition* and *multimedia learning as knowledge construction*. If you view multimedia learning as response strengthening, then multimedia is a feedback delivery system. If you view multimedia learning as information acquisition, then multimedia is an information delivery system. If you view multimedia learning as knowledge construction, then multimedia is a cognitive aid.

Multimedia Learning as Response Strengthening

According to the response strengthening view, learning involves increasing or decreasing the connection between a stimulus and a response. The underlying principle is that the connection is strengthened if a response is followed by reward and is weakened if the response is followed by punishment. This