SECTION 1
Introduction to Multimedia Learning

People learn better from words and pictures than from words alone. This hypothesis is the basis for the program of research described in this book. In short, this book takes a research-based approach to multimedia instruction – that is, instruction that consists of words and pictures rather than words alone. Specifically, the central question addressed in this book is: How can we design multimedia instruction that improves learner understanding of the presented material?

Chapter 1 explores the promise of multimedia learning by offering definitions of key terms and by examining fundamental distinctions that will help you understand research on multimedia learning. Research on multimedia learning draws from and contributes to the science of learning (i.e., the scientific study of how people learn), the science of instruction (i.e., the scientific study of how to help people learn), and the science of assessment (i.e., the scientific study of determining what people have learned). A key distinction is between two goals of multimedia research – to contribute to instructional practice and to contribute to learning theory. The multimedia design principles presented in this book are intended to address both goals and reflect an example of what Stokes (1997, p. 73) calls "use-inspired basic research."

Chapter 2 explores the science of learning by summarizing a research-based theory of how people learn from words and pictures, which I call the cognitive theory of multimedia learning. The theory is based on research in cognitive science, including the ideas of dual channels, limited capacity, and active processing. The cognitive theory of multimedia learning can help you understand how we generated to-be-tested design principles and how we explained when the principles do and do not

apply. I also describe how the theory can be extended to incorporate social, affective, motivational, and metacognitive processes. In short, this chapter helps you see how the instructional design principles described in this book are grounded in theory.

Chapter 3 explores the science of instruction by summarizing the methods we used to test the instructional design principles described in this book. The chapter gives you examples of the multimedia lessons and tests we used, including paper-based annotated illustrations, computer-based narrated animation, computer-based narrated video, computer-based slideshows, and computer-based games and simulations. I also show you how we created experimental comparisons in which we compared the test performance of a group that learned from a multimedia lesson containing a to-be-tested feature versus a group that learned from the lesson without the feature. In short, this chapter helps you see how the instructional design principles described in this book are based on evidence.

Chapter 4 explores the science of assessment by summarizing advances in how to measure learning outcomes and learning processes. Specifically, this chapter examines: (a) how we can know what someone has learned from a multimedia lesson, and (b) how we can know what processing is going on in the learner's information processing system during learning from a multimedia lesson.

Chapter 5 presents the multimedia design principle that started this field – the multimedia principle, which states that people learn better from words and graphics than from words alone.

REFERENCE

Stokes, D. E. (1997). *Pasteur's quadrant: Basic science and technological innovation.* Washington, DC: Brookings Institution Press.

1 The Promise of Multimedia Learning

CHAPTER SUMMARY

Multimedia learning refers to learning from words and pictures. *Multimedia instruction* refers to the presentation of material using both words and pictures, with the intention of promoting learning. The case for multimedia learning rests in the premise that learners can better understand an explanation when it is presented in words and pictures than when it is presented in words alone. Multimedia messages can be characterized based on the *delivery media* (e.g., amplified speaker and computer screen), *presentation mode* (e.g., words and pictures), or *sensory modalities* (e.g., auditory and visual). The design of multimedia instructional messages can be based on a *technology-centered approach* that focuses on the capabilities of advanced technologies, or on a *learner-centered approach* that focuses on the nature of the human cognitive system. Multimedia learning may be viewed as *response strengthening* (in which multimedia environments are used as drill-and-practice systems), *information acquisition* (in which multimedia messages serve as information delivery vehicles), or as *knowledge construction* (in which multimedia messages include aids to sense-making). Three possible learning outcomes are *no learning* (as indicated by poor retention and poor transfer performance), *rote learning* (as indicated by good retention and poor retention performance), and *meaningful learning* (as indicated by good retention and good transfer performance). Meaningful learning outcomes depend on the cognitive activity of the learner during learning rather than on the learner's behavioral activity during learning. The goal of *basic research* is to contribute a theory of learning whereas the goal of *applied research* is to derive principles of instructional design; merging these goals results in basic research on applied issues where the goal is to derive principles of multimedia design that are both grounded in cognitive theory and supported by empirical evidence. The study of multimedia learning draws from and contributes to the *science of learning*, the *science of instruction*, and the *science of assessment*.

CHAPTER OUTLINE

What Is Multimedia Instruction?

The Case for Multimedia Learning

What Is Multimedia Instruction?

People learn better from words and pictures than from words alone. This straight-forward statement can be called the multimedia principle and it is the guiding thesis being tested in this book. In short, I am intrigued by the idea that we can improve people's learning by incorporating effective graphics into verbal material. Does adding graphics to words help people learn better, what makes an effective graphic, and how do people learn from words and pictures? These are the kinds of questions I address in this book through more than 200 experimental tests – questions about what works with multimedia instruction, when it works, and how people learn from multimedia instruction.

Although you may think of multimedia instruction as a product of the digital age, multimedia instruction has a long history dating back to 1657, when John Comenius produced the world's first children's picture book, *Orbis Pictus* (which means "*The World in Pictures*" or "*The World Illustrated*"). Each page contained a black and white line drawing of an aspect of the world ranging from the parts of a house, to elements in a barber shop, to the types of water birds, to the parts of a plant. Each element was numbered, and a corresponding legend gave its name and description in Latin and in the child's native language, such as exemplified in Figure 1.1. As noted in the preface to an English-language version (Comenius, 1887), the book became the most popular textbook in Europe for a century. As you can see, the world's first illustrated school book is based on the multimedia principle, as articulated in the preface (p. vii): "the teaching of words and things must go together." There was no research base to guide Comenius's efforts hundreds of years ago, but as you will see in this book, we are in the process of building evidence-based principles to guide anyone interested in designing multimedia instruction in the twenty-first century.

In today's world, the term *multimedia instruction* means different things to different people. For some people, multimedia instruction involves standing

The Air.   VI.   Aër.

A cool *Air*, 1.   *Aura*, 1.
breatheth gently.   spirat leniter.
   The *Wind*, 2.   *Ventus*, 2.
bloweth strongly.   flat valide.
   A *Storm*, 3.   *Procella*, 3.
throweth down Trees.   sternit Arbores.
   A *Whirl-wind*, 4.   *Turbo*, 4.
turneth it self in a round   agit se in gyrum.
compass.
   A Wind *under Ground*, 5.   Ventus *subterraneus*, 5.
causeth an *Earthquake*.   excitat *Terræ motum*.
   An Earthquake causeth   Terræ motus facit
gapings of the Earth,   Labes (& ruinas.) 6.
(and falls of Houses.) 6.

Figure 1.1 Page from *Orbis Pictus*

in a room that has screens displaying still images, videos, and/or animations and speakers transmitting music, sounds, and/or narration. Others may think of walking around in immersive virtual reality playing the latest digital game. Perhaps it makes you think of sitting at a desktop computer that is playing a narrated video or a narrated animation. Maybe you think of looking up a topic

via an online encyclopedia or news site on your cellphone or tablet. Another example of multimedia instruction is a slideshow lecture in which someone presents slides from a computer projected onto a screen and talks about each one. Even low-tech environments allow for multimedia instruction, such as a "chalk and talk" presentation in which an instructor writes or draws on a blackboard (or uses an overhead projector) while presenting a lecture. Finally, the most basic form of multimedia instruction is a textbook lesson consisting of printed text and illustrations.

I define multimedia instruction as the presentation of material using both words and pictures, with the intention of promoting learning. By words, I mean that the material is presented in *verbal form*, such as using printed or spoken text. By pictures (which also can be called graphics), I mean that the material is presented in *pictorial form*, including using static graphics such as illustrations, graphs, photos, and maps, or dynamic graphics such as animations and videos. This definition is broad enough to cover each of the multimedia scenarios I just described in the previous paragraph, ranging from online encyclopedia entries and textbook lessons to slideshow lectures and online narrated videos to digital games and simulations in virtual reality. For example, in a multimedia encyclopedia the words can be presented as onscreen text or as narration and the pictures can be presented as graphics or animations. In a textbook, the words can be presented as printed text and the pictures as illustrations (or other kinds of graphics).

For purposes of conducting research, I have focused the definition of multimedia instruction on just two presentation formats – words and pictures. I have opted to limit the definition to just two formats – verbal and pictorial – because the research base in cognitive science is most relevant to this distinction. Thus, what I call multimedia learning is more accurately called dual-mode, dual-format, dual-code, or dual-channel learning.

Is multimedia a noun or an adjective? When used as a noun, *multimedia* refers to a technology for presenting material in both visual and verbal forms. In this sense, multimedia means multimedia technology – devices used to present visual and verbal material. When used as an adjective, multimedia can be used in the following contexts:

multimedia learning – learning from words and pictures

multimedia instruction (or *multimedia instructional message* or *multimedia instructional presentation*) – presentations involving words and pictures that are intended to foster learning

multimedia message or *multimedia presentation* – presentations involving words and pictures

My focus in this book is on the design of multimedia instructional messages that promote multimedia learning.

In the remainder of this chapter, I present the case for multimedia learning, and then I examine three views of multimedia messages, two approaches to multimedia design, three metaphors of multimedia learning, three kinds of multimedia learning outcomes, two kinds of active learning, two goals of multimedia research, and three essential elements in multimedia research.

The Case for Multimedia Learning

The case for multimedia learning is based on the idea that instructional messages should be designed in light of how the human mind works. Let's assume that humans have two information processing systems, one for verbal material and one for visual material. 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 our full capacity 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 explanations are the quantitative rationale and the qualitative rationale. The quantitative rationale is that more material can be presented in two channels than on one channel – just as more traffic can travel over two lanes than one lane. In the case of explaining how a car's braking system 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 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 occurs when learners are able to mentally integrate corresponding pictorial 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 formal 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 1000 words (or any number of words), but it can help us understand the words.

The most intriguing aspect of the qualitative rationale is that understanding occurs when learners are able to build meaningful connections between pictorial and verbal representations, such as being able to see how the words "the piston moves forward in the master cylinder" relates to the forward motion of a piston in the master cylinder in an animation of a car's braking system. 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 cognitive theory of multimedia learning that is described in Chapter 2.

An instructional message is a communication that is intended to foster learning. In presenting an instructional message to learners, instructional designers have two main formats available – words and pictures. Words include speech and printed text; pictures include static graphics (such as illustrations and photos) and dynamic graphics (such as animations and videos). For hundreds of years, the major format for presenting instructional messages has been words, including lectures and books. In short, verbal modes of presentation have dominated the way we convey explanations to one another and verbal learning has dominated education. Similarly, verbal learning has been the major focus of educational research.

The advent of computer technology has enabled an explosion in the availability of visual ways of presenting material, including large libraries of static images, and the ability to create and edit compelling dynamic images in the form of animations and videos. In light of the power of computer graphics, it may be useful to ask whether it is time to expand instructional messages beyond the purely verbal. What are the consequences of adding pictures to words? What happens when instructional messages involve both verbal and visual modes of learning? What affects the way that people learn from words and pictures? In short, how can multimedia presentations foster meaningful learning? These are the kinds of questions addressed in this book.

Three Views of Multimedia Messages

The term *multimedia* can be viewed in three ways – based on the devices used to deliver an instructional message (i.e., the delivery media), the representational formats used to present the instructional message (i.e., the presentation modes), or the sense modalities the learner uses to receive the instructional message (i.e., sensory modalities).

The delivery media view. The most obvious view is that multimedia means the presentation of material using two or more delivery devices. The focus is on the physical system used to deliver the information, such as computer screens, amplified speakers, projectors, video recorders, blackboards, and human voice boxes. For example, in computer-based multimedia, material can be presented via the screen and via the speakers. These devices can even be broken down further by defining each window on a computer screen as a separate delivery device and each sound track coming from a speaker as a separate delivery device. In lecture-based multimedia, material can be presented via a projector onto a screen and via the lecturer's voice. In the strictest interpretation of the delivery media view, a textbook does not constitute multimedia because the only presentation device is ink printed on paper. Similarly, an e-book does not involve multimedia because all the information is coming from one source – your device's screen.

What's wrong with the delivery media view of multimedia? Technically, it is the most accurate view because it focuses on the media used to present information, but psychologically, it does more to confuse the issue than to clarify it. The focus is on the devices used to present information rather than on how people learn – that is, the focus is on technology rather than on learners. Therefore, I do not take the delivery media view in this book.

The presentation modes view. A second view is that multimedia means the presentation of material using two or more presentation modes. The focus is on the way that material is represented, such as through the use of words or pictures. For example, in computer-based multimedia, material can be presented verbally as onscreen text or narration and pictorially as static graphics or animation or video. In lecture-based multimedia, material can be presented verbally as speech and pictorially as projected graphics or video. In a textbook, material can be presented verbally as printed text and pictorially as static graphics.

This view is consistent with a learner-centered approach if we assume that learners are able to use various coding systems to represent knowledge, such as verbal and pictorial knowledge representations. Although conventional wisdom is that a picture can be converted into words and vice versa, research on mental representations suggests that verbal ways of representing knowledge may be qualitatively different from pictorial ways of representing knowledge. In short, the presentation modes view of multimedia is consistent with a cognitive theory of learning, which assumes humans have separate information processing channels for verbal and pictorial knowledge. Paivio's (1986, 2007; Sadoski & Paivio, 2013) dual-code theory presents the most coherent theoretical and empirical evidence for this idea.

The sensory modality view. The third view, while also consistent with a learner-centered approach, takes a somewhat different approach. According to

the sensory modalities view, multimedia means that two or more sensory systems are involved in the learner. Instead of focusing on codes used to represent knowledge in the learner's information processing system, the sensory modalities view focuses on the sensory receptor the learner uses to perceive the incoming material, such as the eyes and the ears. For example, in a computer-based environment an animation can be presented visually and a narration can be presented auditorily. In a lecture scenario, the speaker's voice is processed in the auditory channel and the slides from the projector are processed in the visual channel. In a textbook, illustrations and printed text are both processed visually, at least initially.

This view is learner-centered because it takes the learner's information processing activity into account. Unlike the presentation modes view, however, the sensory modalities view is that multimedia involves presenting material that is processed visually and auditorily. This distinction is based on the idea that humans process visual images and sounds in qualitatively different ways. In short, the sensory modalities view of multimedia is consistent with a cognitive theory of learning, which assumes humans have separate information processing channels for auditory and visual processing. Baddeley's (2019; Baddeley & Eysenck, 2015) model of working memory presents the most coherent theoretical and empirical evidence for this idea.

Table 1.1 summarizes the differences among these three views. In sum, I reject the delivery media view because it emphasizes the delivery technology over the learner. Both the presentation modes and sensory modalities views focus on the information processing system of the learner and assume that humans process information in more than one channel, a proposal that I call the dual-channel assumption. However, they differ in the way they conceptualize the nature of the two channels: the presentation modes view distinguishes between separate systems for processing verbal and pictorial knowledge, whereas the sensory modes view distinguishes between separate systems for auditory and visual processing (i.e., for processing sounds and visual images). Although my definition of

Table 1.1 Three views of multimedia

View	Definition	Examples
Delivery media	Two or more delivery devices	Computer screen and amplified speakers; projector and lecturer's voice
Presentation mode	Verbal and pictorial representations	Onscreen text and animation; printed text and illustrations
Sensory modality	Auditory and visual senses	Narration and animation; lecture and slides