Handbook of Research on Learning and Instruction

Richard E. Mayer, Patricia A. Alexander

Instruction Based on Visualizations

Publication details
Richard E. Mayer
Published online on: 14 Oct 2016

How to cite :- Richard E. Mayer. 14 Oct 2016, Instruction Based on Visualizations from: Handbook of Research on Learning and Instruction Routledge
Accessed on: 22 Nov 2023

PLEASE SCROLL DOWN FOR DOCUMENT

Full terms and conditions of use: https://www.routledgehandbooks.com/legal-notices/terms

This Document PDF may be used for research, teaching and private study purposes. Any substantial or systematic reproductions, re-distribution, re-selling, loan or sub-licensing, systematic supply or distribution in any form to anyone is expressly forbidden.

The publisher does not give any warranty express or implied or make any representation that the contents will be complete or accurate or up to date. The publisher shall not be liable for an loss, actions, claims, proceedings, demand or costs or damages whatsoever or howsoever caused arising directly or indirectly in connection with or arising out of the use of this material.
INTRODUCTION

People learn better from words and pictures than from words alone. This proposal, which can be called the multimedia instruction hypothesis, is examined in this chapter. For thousands of years, the main medium of instruction has involved the use of words—progressing from words in spoken form to written form to printed form to electronic form. Recently, advances in computer-based visualization technology have enabled the incorporation of sophisticated graphics in instruction—including animation, video, illustrations, and photos. In this chapter, I explore the question of whether adding visualizations to words in instructional messages can improve student learning.

Multimedia instruction occurs when instructional messages contain both words and pictures (Mayer, 2009). An instruction message is a communication intended to promote learning, whereas a multimedia instruction message is a communication that contains both words and pictures, and is intended to promote learning (Mayer, 2009). As summarized in Table 22.1, words are verbal representations, such as printed text (delivered on a page or screen) or spoken text (delivered face-to-face or via speakers); and pictures are visual-spatial representations, such as static graphics (including illustrations, drawings, photos, maps, diagrams, charts, figures, and tables delivered on a page or screen) or dynamic graphics (including animation and video delivered on a screen). In this chapter, I use the terms pictures, graphics, and visualizations interchangeably, although in some venues visualizations refer only to computer-rendered visual-spatial representations.

An instructional visualization (or instructional picture or instructional graphic) is a visual-spatial representation intended to promote learning. Instructional visualizations can vary along several dimensions:

- **Realism**—pictures can vary from high realism (e.g., a photo or video) to low realism (e.g., a line drawing or an animated line drawing),
- **Dynamism**—pictures can be static (e.g., a drawing or photo) or dynamic (e.g., an animation or video),
Interactivity — pictures can be interactive (e.g., a series of drawings that can be paced by the learner or an animation that can be stopped and started by the learner) or non-interactive (e.g., a drawing or continuous animation),

Dimensionality — pictures can be presented in 2D or 3D form,

Visual-spatial character — pictures can be visual representations (e.g., a drawing or photo of an object) or spatial representations (e.g., a chart or table or maps), and

Delivery medium — pictures can be presented on a page or screen.

In most of the research presented in this chapter, I focus on pictures that are low in realism, non-interactive, visual, two-dimensional, and that can be either static or dynamic, and that can be delivered on a page or screen.

A BRIEF HISTORY OF INSTRUCTIONAL VISUALIZATIONS

There have been three major phases in the technology supporting instructional visualizations: books, film, and computers.

Books

A major breakthrough in multimedia instruction occurred more than 350 years ago, when the Czech educator, John Amos Comenius, published a children’s illustrated book entitled, Orbis Pictus (The World in Pictures) in 1658. As exemplified in Figure 22.1 (based on an 1887 English-language version), each page contained a line drawing of a scene with numbered objects along with accompanying text that named and explicated each object in both the reader’s first language and in Latin. Comenius’s book was a sort of illustrated encyclopedia containing what he called “nomenclature and pictures of all the chief things that are in the world” (Comenius, 1887, p. xi). Orbis Pictus is widely recognized as the first instructional book to combine words and pictures, and stands as “the most popular illustrated textbook ever written for children” (Saettler, 2004, p. 31). Thus, Orbis Pictus is the forerunner of modern illustrated textbooks in particular and multimedia instruction in general.

In his preface, Comenius offered a theoretical rationale for why words and pictures should be learned together: “There is nothing in our understanding that was not before in the sense” (Comenius, 1887, p. xiv). He objected to the words-only approach to instruction on the grounds that when “things which are to be learned are offered to scholars without being understood or being rightly presented to the senses, it cometh to pass, that the work of teaching and learning goeth heavily onward and affordeth little benefit” (Comenius, 1887, p xiv). In short, the rationale for multimedia instruction
is that learners’ understanding of text depends on their being able to relate the words to corresponding concrete visual representations and situations.

**Motion Pictures and Television**

Another milestone in multimedia instruction occurred approximately 100 years ago (1911–1914), when Thomas Edison released the world’s first instructional films for classroom showing, with titles in history such as *The Minute Men* and in science such as *Life History of the Silkworm* (Saettler, 2004). Edison predicted that “the motion picture is destined to revolutionize our educational system” (Cuban, 1986, p. 9), and indeed by 1931, 25 states had bureaus dedicated to visual education using motion pictures.
(Saettler, 2004). In 1954, the first educational television station (KUHT) began broadcasting in Houston, Texas, and by the 1970s the Children’s Television Workshop created landmark educational television programming such as *Sesame Street* and *The Electric Company*.

**Computers**

Beginning in the 1960s, widespread use of computer-based instructional systems became feasible (Cuban, 1986), and within the past decades, we have witnessed important advances in communication technology, including the widespread availability of the Internet, and in graphics technology, including affordable graphics software for producing educational illustrations, animation, and video. Today, it is possible to create compelling computer-based visual simulations on laptop computers, to deliver stunning graphics via hand-held devices, and to offer the experience of immersive virtual reality (Mayer, 2014a). An important educational issue concerns how best to use the graphics capabilities in book-, film-, and computer-based technologies to improve student learning.

**THEORETICAL FRAMEWORK**

**Does Multimedia Instruction Work?**

It is important to determine whether there is any value added to student learning outcomes by adding graphics to words. In short, a fundamental research question is: Does adding graphics to words help people learn better than presenting words alone? For example, if someone looks up “brakes” in an online encyclopedia or wiki, they may come across a section on hydraulic brakes that contains the following explanation of how brakes work:

> When the driver steps on the car’s brake pedal, a piston moves forward inside the master cylinder. The piston forces brake fluid out of the master cylinder and through the tubes to the wheel cylinders. In the wheel cylinders, the increase in fluid pressure makes a smaller set of pistons move. These smaller pistons activate the brake shoes. When the brake shoes press against the drum both the drum and wheel stop or slow down.

This explanation provides a step-by-step description of the causal chain in which a change in state in one part (e.g., piston moves forward in master cylinder) causes a change in state in another part (e.g., brake fluid is forced through the tubes) which causes a change in state in another part (e.g., smaller set of pistons move) and so on. The text is intended to help the learner build a *causal model* of the braking system, consisting of each component, the changes in each component, and the relations among the changes in the components.

How well do people learn the explanation from printed words? On an immediate retention test in which they were asked to write all they can remember about how brakes work, people generated less than 25% of the important information (Mayer, 2009). On an immediate transfer test in which they were asked to generate as many answers as possible to open-ended questions (e.g., “Suppose you press the brake pedal in your car but the brakes don’t work. What could have gone wrong?”), people
averaged less than one acceptable answer per question (Mayer, 2009). In short, people do not appear to learn much from reading a text that explains how something works. Similar results were obtained when people listened to an explanation rather than read it (Mayer, 2009).

What can be done to improve people’s understanding of verbal explanations? We can add a series of frames consisting of line drawings that depict the actions described in the verbal explanation. Figure 22.2 shows line drawings depicting the braking system before and after the driver steps on the car’s brake pedal along with accompanying printed text. Alternatively, we can add animation (based on line drawings) to a narration. As you can see, the added visualizations provide a concrete context for making sense of the words.

Overall, across 11 experimental comparisons involving lessons on topics such as how brakes, pumps, or lightning works, people performed better on transfer tests when they learned from printed text and illustrations than from printed text alone (Mayer, 1989, Experiments 1 and 2; Mayer, Bove, Bryman, Mars, & Tapango, 1996, Experiment 2; Mayer & Gallini, 1990, Experiments 1, 2, and 3) or from narration and animation than from narration alone (Mayer & Anderson, 1991, Experiment 1; Mayer & Anderson, 1992, Experiments 1 and 2; Moreno & Mayer, 1999, Experiment 1; Moreno & Mayer, 2002a, Experiment 1). The median effect size favoring words and pictures over words alone is $d = 1.4$, which is considered a large effect. These results are evidence for the multimedia principle: People learn better from words and pictures than from words alone (Mayer, 2009).

An important consideration is whether the multimedia principle applies in classroom learning situations. For example, a reasonable criticism of the foregoing studies...
is that they were mainly conducted in lab settings and used very short lessons. In a set of three studies, students who learned about learning principles from a lecture followed by a video showing a case example performed better on a transfer test than did students who learned from a lecture followed an equivalent text booklet describing a case example (Moreno & Ortegano-Layne, 2008, Experiment 1; Moreno & Valdez, 2007, Experiments 1 and 2). In two studies, beginning trade apprentices and trainees in a work-related training program performed better on subsequent transfer tests if their training materials consisted of text and printed diagrams rather than text alone, or audio narration and printed diagrams rather than diagrams alone (Kalyuga, Chandler, & Sweller, 1998, Experiment 1; Kalyuga, Chandler, & Sweller, 2000, Experiment 1). Thus, there is promising evidence that the multimedia principle may extend to more authentic learning environments.

In a recent review, Butcher (2014, p. 198) concluded that the multimedia principle is based on “robust research findings,” but also noted that the effectiveness of combining words and pictures may depend on the some boundary conditions. In particular, the learner’s prior knowledge may serve as a boundary condition for the multimedia principle in which adding pictures to words is particularly helpful for low-knowledge learners but not for high-knowledge learners. For example, Mayer and Gallini (1990) asked students to read a booklet explaining how brakes, pumps, or electrical generators work and then take a transfer test. For students who reported low levels of prior mechanical knowledge, adding line drawings to the printed text greatly improved transfer test performance as compared to presenting text alone. In contrast, for students who reported high levels of prior mechanical knowledge, adding diagrams to printed text did not greatly improve transfer test performance.

In another set of experiments, Kalyuga, Chandler, and Sweller (1998, 2000) taught students how to solve practical engineering problems using diagrams accompanied by printed text or audio narration. Beginning trainees learned better from words and pictures than from words or pictures alone, but as trainees gained more experience they learned worse from words and pictures than from words or pictures alone. Kalyuga (2014) refers to this pattern as the expertise reversal effect: Instructional methods that improve learning for low-knowledge learners may be ineffective or even harmful for high-knowledge learners. It appears that domain-specific prior knowledge may be a useful variable to consider when designing instruction involving visualizations. In particular, adding pictures to words may be particularly helpful for low-knowledge learners, presumably because they are less able to create and link images to words on their own.

Another individual differences dimension that has received much attention in education is cognitive style—such as the distinction between visualizers and verbalizers (Massa & Mayer, 2006). In particular, a common claim is that visualizers would benefit more from adding pictures to words than would verbalizers (Pashler, McDaniel, Royer, & Bjork, 2008). However, in a systematic set of studies, Massa and Mayer (2006) gave students a computer-based lesson on electronics that consisted mainly of text along with help frames that combined mainly text or mainly pictorial material. Both visualizers and verbalizers performed better on subsequent transfer tests if they had received pictorial help frames rather than text help frames. Overall, Pashler, McDaniel, Royer, and Bjork (2008) conclude that there is not sufficient evidence to provide visual instruction to visualizers and verbal instruction to verbalizers. Thus, it appears that cognitive style is not a major individual differences variable for teaching with visualizations.
How Does Multimedia Instruction Work?

In the previous section, I examined some evidence showing that adding visualizations to a word-based lesson can improve students’ understanding of the material. In this section, my goal is to examine how the additional visualizations affect the process of learning. I begin with three principles from cognitive science concerning how learning works—dual channels, limited capacity, and active processing (Mayer, 2009, 2014b). The dual channels principle is that people have separate channels for processing words and pictures (Baddeley, 1999; Paivio, 1986, 2001). The limited capacity principle is that people are able to engage in only a limited amount of cognitive processing in each channel at any one time (Baddeley, 1999; Sweller, Ayres, & Kalyuga, 2011). The active processing principle is that meaningful learning occurs when people engage in appropriate cognitive processing during learning, including selecting relevant incoming words and pictures for further processing, organizing the selected words and pictures into coherent mental representations, and integrating the representations with each other and with relevant prior knowledge activated from long-term memory (Fiorella & Mayer, 2015; Mayer, 2009; Wittrock, 1989).

Figure 22.3 summarizes a cognitive model of how multimedia instruction works. There are two channels—a verbal channel across the top row for processing words and verbal representations, and a pictorial channel across the bottom row for processing pictures and pictorial representations. The three boxes in the columns of the model represent sensory memory, working memory, and long-term memory. Sensory memory briefly holds incoming sounds and images in sensory form with unlimited capacity. Working memory can be used to temporarily hold and manipulate selected verbal and pictorial representations with limited capacity. Long-term memory is the learner’s permanent storehouse of knowledge with unlimited capacity.

The arrows in the model represent cognitive processes during learning required for meaningful learning—selecting, organizing, and integrating. Selecting words refers to paying attention to some of the incoming spoken words that are fleeting in auditory sensory memory, thereby transferring them to working memory for further processing in the verbal channel. Selecting images refers to paying attention to some of the incoming pictures and printed words that are fleeting in visual sensory memory, and transferring them to working memory for further processing in the pictorial channel. Printed words are initially processed in the pictorial channel in working memory and then shifted to the verbal channel in working memory (indicated by the arrow from images to sounds). Organizing words refers to building a coherent verbal representation in the verbal channel of working memory, and organizing images refers to building a coherent pictorial representation in the pictorial channel of working memory.

Figure 22.3 A cognitive model of learning with words and pictures.
Integrating refers to building connections between the verbal and pictorial representations and with relevant prior knowledge activated from long-term memory.

What happens when a learner is in an instructional situation involving words and pictures (e.g., reading an illustrated text, viewing a narrated animation, or interacting with a multimedia instructional simulation)? If the instruction is well designed and appropriate for the learner, we can expect that all five cognitive processes will be activated, resulting in a meaningful learning outcome, which is then stored in long-term memory. The act of integrating verbal and pictorial representations with each other is an important step in promoting deep understanding.

In contrast, consider what happens when the learner receives only words (e.g., printed or spoken text). In this case, inexperienced learners may not be able to generate relevant pictorial representations on their own, so they do not have the opportunity to build a pictorial representation (i.e., indicated by the organizing images arrow) or to integrate verbal and pictorial representations (i.e., indicated by the integrating arrow within working memory). Thus, the learning outcome for words-only instruction is not as well developed as for multimedia instruction.

In learning with words and pictures, learners may experience three kinds of demands on their limited processing capacity in working memory: extraneous processing, essential processing, and generative processing (Fiorella & Mayer, 2015; Mayer, 2009, 2014b, Mayer & Moreno, 2003; related to similar concepts in cognitive load theory by Paas & Sweller, 2014, and Sweller, Ayres, & Kalyuga, 2011). Extraneous processing refers to cognitive processing that does not support the learning objective and can be caused by poor layout. For example, placing the illustration on one page and the corresponding text on another page causes the learner to have to scan back and forth, wasting precious cognitive processing capacity. It follows that an important instructional goal is to reduce extraneous processing, so that the learner has sufficient remaining capacity to engage in essential and generative processing.

Essential processing refers to cognitive processing that is required to mentally represent the material and is caused by the inherent complexity of the material for the learner. Essential processing requires the process of selecting in Figure 22.3, as well as some initial amount of organizing. Even if we could eliminate extraneous processing, the demands of essential processing might overwhelm the learner’s cognitive system. Thus, an important instructional goal is to manage essential processing in a way that prevents cognitive overload.

Generative processing refers to cognitive processing aimed at making sense of the material and is caused by the learner’s motivation. Generative processing requires the processes of integrating and organizing in Figure 22.3. Even if we reduce extraneous processing and manage essential processing so processing capacity is available, learners may not use that available capacity to engage in generative processing. Thus, an important instructional goal is to foster generative processing.

In short, based on the model of multimedia instruction presented in Figure 22.3, three kinds of threats to appropriate cognitive processing during learning: too much extraneous processing (so a goal is to reduce extraneous processing), too much essential processing (so a goal is to manage essential processing), and too little generative processing (so a goal is to foster generative processing).

Two more recent enhancements to model of multimedia instruction presented in Figure 22.3 involve the role of motivation and metacognition (Fiorella & Mayer, 2015; Mayer, 2011). Motivation refers to the processes that initiate and maintain the learner’s effort to engage in the cognitive processes of learning. Metacognition refers to learners
selecting their learning processes, monitoring their effectiveness, and adjusting them as needed. Motivation and metacognition can be represented as arrows from long-term memory back to the other arrows, but much more work is needed to specify the mechanisms by which motivational and metacognitive processes affect multimedia learning.

CURRENT TRENDS AND ISSUES

When Does Multimedia Instruction Work?

Although learning from words and pictures can result in deeper learning than learning from words alone, not all pictorial visualizations are equally effective. In this section, I examine how to design effective visualizations for meaningful learning, based on the three goals described in the previous section. In short, this section briefly examines instructional design principles for reducing extraneous processing, managing essential processing, and fostering generative processing.

As shown in Table 22.2, principles for reducing extraneous processing include the coherence principle, the signaling principle, the redundancy principle, the spatial contiguity principle, and the temporal contiguity principle (Mayer & Fiorella, 2014). The coherence principle is that people learn better from multimedia lessons when extraneous words and pictures are excluded rather than included. For example, a seductive detail is interesting but irrelevant verbal information, such as verbal descriptions of airplanes or people being struck by lightning in a lesson on how lightning storms develop. Similarly, pictorial seductive details are interesting but irrelevant visualizations, such as interspersed pictures or video of lightning storms in a lesson on how lightning storms develop. In nine experimental tests, adding seductive details to an illustrated text (Harp & Mayer, 1997, Experiment 1; Harp & Mayer, 1998, Experiments 1, 2, 3, and 4) or a narrated animation (Mayer, Heiser, & Lonn, 2001, Experiment 3) on lightning formation or adding highly interesting sentences to a multimedia lesson on cold viruses (Mayer, Griffith, Naftaly, & Rothman, 2008, Experiment 1) or digestion (Mayer, Griffith, Naftaly, & Rothman, 2008, Experiment 2) or adding highly interesting but irrelevant graphics to a multimedia lesson on distance learning (Sung & Mayer, 2012) resulted in lower performance on transfer tests, with a median effect size of \( d = 1.3 \) favoring the more concise lesson.

Similarly, in six experimental comparisons, students performed better on transfer tests if extraneous words had been eliminated from lessons containing printed text and illustrations on ocean waves (Mayer & Jackson, 2005, Experiments 1a and 1b) or lightning (Mayer, Bove, Bryman, Mars, & Tapango, 1996, Experiments 1, 2, and

<table>
<thead>
<tr>
<th>Table 22.2 Five Principles for Reducing Extraneous Processing</th>
</tr>
</thead>
<tbody>
<tr>
<td>Principle</td>
</tr>
<tr>
<td>-------------------</td>
</tr>
<tr>
<td>Coherence</td>
</tr>
<tr>
<td>Signaling</td>
</tr>
<tr>
<td>Redundancy</td>
</tr>
<tr>
<td>Spatial contiguity</td>
</tr>
<tr>
<td>Temporal contiguity</td>
</tr>
</tbody>
</table>

Note. \( n \) = number of comparisons; \( d \) = median effect size.
3), or lessons containing narration and animation on ocean waves (Mayer & Jackson, 2005, Experiment 2), yielding a median effect size of $d = 0.8$. Finally, students performed better on transfer tests when background music was deleted from computer-based narration and animation on lightning (Moreno & Mayer, 2000a, Experiment 1) or brakes (Moreno & Mayer, 2000a, Experiment 2), with a median effect size of $d = 1.1$.

Overall, as summarized in the first line of Table 22.2, across 17 experiments carried out in our lab, the median effect size favoring the coherence principle was $d = 1.0$, which is a large effect. When extraneous material is eliminated, people can focus their cognitive processing on the essential words and pictures in the lesson.

The signaling principle is that people learn better from multimedia lessons when the essential words are highlighted. Signaling can involve adding voice emphasis to essential words, adding an outline, adding headings, or adding a graphic organizer. In this way, the learner may be better able to make connections between the essential verbal material and the corresponding portions of the visualization. Modest preliminary support for the signaling principle can be seen across six experimental comparisons carried out in our lab involving a computer-based multimedia lesson on how airplanes achieve lift (Mautone & Mayer, 2001, Experiments 3a and 3b) and paper-based lessons on lightning (Harp & Mayer, 1998, Experiment 3a) and biology (Stull & Mayer, 2007, Experiments 1, 2, and 3), yielding a median effect size of $d = 0.5$. In short, as summarized in the second row of Table 22.2, adding signals resulted in a median improvement of 0.5 standard deviation in transfer test performance. Signaling may be more effective when the display is complex (Jueng, Chandler, & Sweller, 1997) and when it used sparingly (Stull & Mayer, 2007).

The redundancy principle is that people learn better from onscreen visualizations with narration than from onscreen visualizations with narration and onscreen text. In short, the redundancy principle advises against adding onscreen text to a narrated animation or narrated still graphics. The redundancy principle is supported across five experimental comparisons carried out in our lab involving computer-based lessons on lightning (Mayer, Heiser, & Lonn, 2001, Experiments 1 and 2; Moreno & Mayer, 2002a, Experiment 2) and computer-based simulation games in botany (Moreno & Mayer, 2002a, Experiments 2a and 2b), yielding a median effect size of $d = 0.7$. The third line of Table 22.2 summarizes this finding that transfer test performance was improved by an average of 0.7 standard deviations when redundant onscreen text was removed from a multimedia lesson. When people pay attention to the onscreen text, they are not able to attend to the visualization, and they may waste cognitive capacity by trying to reconcile the two verbal streams. Some important boundary conditions for the redundancy principle are that onscreen text can be helpful when the words are unfamiliar to the learner, the learner is not a native speaker of the language, the learner has hearing difficulties, the onscreen text is short and placed next to the corresponding part of the graphic, or the onscreen text rewords the audio (Mayer, 2009; Mayer & Johnson, 2008; Yue, Bjork, & Bjork, 2013).

The spatial contiguity principle is that people learn better from multimedia lessons when printed words are placed near rather than far from the corresponding part of the graphic on the screen or page. By placing corresponding words and graphics near each other, the learner is less likely to have to scan around the page or screen trying to figure out where to look. Strong and consistent support for the spatial contiguity principle can be found across eight experimental comparisons involving paper-based lessons on lighting (Mayer, Steinhoff, Bower, & Mars, 1995, Experiments 1, 2, and 3) and brakes...
(Mayer, 1989, Experiment 2), and computer-based lessons on lightning (Moreno & Mayer, 1999, Experiment 1) and brakes (Johnson & Mayer, 2012, Experiments 1, 2, and 3), yielding a median effect size of $d = 1.0$. The fourth row of Table 22.2 shows that students score an average of 1 standard deviation better on transfer tests when printed words are placed next to the part of graphic they refer to. In reviews of research on the spatial contiguity effect, Ginns (2006) and Ayres and Sweller (2014) have reported strong supporting evidence, but there are some possible boundary conditions (Mayer, 2009). Specifically, the spatial contiguity principle may apply most strongly for low-knowledge learners (Kalyuga, 2014), when the material is complicated (Ayres & Sweller, 2005), and when the learner places the words next to graphics through interactivity (Bode-mer, Ploetzner, Feuerlein, & Spada, 2004).

The temporal contiguity principle is that people learn better when pictures and narration are presented simultaneously rather than successively. Learners are better able to make connections between words and pictures when the spoken words correspond to the visualizations being presented on the screen. As shown in the fifth line of Table 22.2, across eight experimental comparisons carried out in our lab, transfer test performance is generally better for simultaneous rather than successive presentation, yielding a median effect size of $d = 1.3$. The evidence comes from computer-based lessons on pumps (Mayer & Anderson, 1991, Experiments 1 and 2a; Mayer & Anderson, 1992, Experiment 1; Mayer & Sims, 1994, Experiment 1), brakes (Mayer & Anderson, 1992, Experiment 2; Mayer, Moreno, Boire, & Vagge, 1999, Experiment 2), lungs (Mayer & Sims, 1994, Experiment 2), and lightning (Mayer, Moreno, Boire, & Vagge, 1999, Experiment 1). Similarly, Ginns (2006) identified 13 experimental comparisons testing the temporal contiguity principle, yielding a median effect size of $d = 0.9$. Some possible boundary conditions are that the effect is eliminated when learners can control the pace and order of presentation (Michas & Berry, 2000) and when the segments are very short (Mayer, Moreno, Boire, & Vagge, 1999; Moreno & Mayer, 1999; Schuler, Scheiter, Rummer, & Gerjets, 2012).

Table 22.3 shows three principles for managing essential processing: the segmenting principle, the pretraining principle, and the modality principle. The segmenting principle is that people learn better from a multimedia lesson when the lesson is broken down into learner-paced segments. The rationale is that when the material is presented in bite-size chunks, the learner is able to completely process the words and pictures in one segment before moving on the next. There is consistent support for the segmenting principle across three experimental comparisons carried out in our lab, including computer-based multimedia lessons on lightning (Mayer & Chandler, 2001, Experiment 2), electric motors (Mayer, Dow, & Mayer, 2003, Experiments 2a and 2b), and geography (Mautone & Mayer, 2007, Experiment 2).

### Table 22.3 Three Principles for Managing Essential Processing

<table>
<thead>
<tr>
<th>Principle</th>
<th>Description</th>
<th>$n$</th>
<th>$d$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Segmenting</td>
<td>Break a continuous lesson into manageable parts.</td>
<td>3</td>
<td>1.0</td>
</tr>
<tr>
<td>Pretraining</td>
<td>Provide pretraining in names and characteristics of each main concept.</td>
<td>6</td>
<td>0.8</td>
</tr>
<tr>
<td>Modality</td>
<td>Accompany graphics with spoken words rather than printed words.</td>
<td>17</td>
<td>1.0</td>
</tr>
</tbody>
</table>

Note. $n =$ number of comparisons; $d =$ median effect size.
Table 22.3, students performed better on transfer tests if they received segmented lessons in which they could control the pacing of segments rather than lessons in which all the material was presented at the same time or continuously, yielding a median effect size of $d = 1.00$. Some possible boundary conditions are that segmenting may be most effective for lower-achieving students rather than higher achieving students (Ayres, 2006) and for learners with lower working memory capacity rather than higher working memory capacity (Lusk et al., 2009).

The pretraining principle is that people learn better from multimedia lessons when they receive pretraining in the names and characteristics of the key concepts. The rationale is that when visualizations are presented along with a verbal explanation, people are better able to make connections between the words and visualizations when they know the names and characteristics of the key elements in the verbal explanation and the visualizations. The pretraining principle is supported across six experimental comparisons carried out in our lab, involving computer-based presentations on brakes (Mayer, Mathias, & Wetzell, 2002, Experiments 1 and 2), pumps (Mayer, Mathias, & Wetzell, 2002, Experiment 3), and computer-based simulation games in geology (Mayer, Mautone, & Prothero, 2002, Experiments 2 and 3) and electronics (Fiorella & Mayer, 2012, Experiment 1), yielding a median effect size of $d = 0.8$. The second line of Table 22.3 shows that transfer test performance was improved substantially when learners received pretraining. A potential boundary condition is that pretraining may be more effective for learners with low rather than high prior knowledge (Pollock, Chandler, & Sweller, 2002).

The modality principle is that people learn better from visualizations with spoken words than from visualizations with printed words. When people pay attention to printed words, they are not paying attention to the visualization, so using spoken text allows the learner to free up capacity in the visual channel. The modality principle was supported across 17 experimental comparisons carried out in conjunction with our lab, involving computer-based multimedia lessons with recorded voice or onscreen text (Harskamp, Mayer, Suhre, & Jansma, 2007, Experiments 1 and 2a; Mayer, Dow, & Mayer, 2003, Experiment 1; Mayer & Moreno, 1998, Experiments 1 and 2; Moreno & Mayer, 1999, Experiments 1 and 2) and computer-based multimedia simulation games with spoken or onscreen text (Moreno & Mayer, 2002b, Experiments 1a, 1b, 1c, 2a, and 2b; Moreno, Mayer, Spires, & Lester, 2001, Experiments 4a, 4b, 5a, and 5b; O’Neil, Mayer, Herl, Thurman, & Olin, 2000, Experiment 1). As shown in the third line of Table 22.3, we found generally strong and consistent support for the modality principle, yielding a median effect size of $d = 1.0$. According to research reviews, the modality principle has received the most research support of any of the principles (Ginns, 2005; Low & Sweller, 2014; Mayer & Pillegard, 2014), but it is important to note that the modality principle is strongest when the material is complex for the learner (Tindale-Ford, Chandler, & Sweller, 1997) and when the pace is fast and not under learner control (Tabbers, Martens, & van Merrienboer, 2004). Spoken words may be ineffective when the message is long or contains unfamiliar words or symbols (Mayer & Pillegard, 2014).

Table 22.4 summarizes principles for fostering generative processing, including the personalization principle, the voice principle, and the embodiment principle. The theoretical rationale is that social cues can create a sense of presence that encourages the learner to try harder to make sense of what the instructor is saying (Mayer, 2014c). The personalization principle is that people learn better when words in a multimedia lesson are presented in conversational style rather than formal style. The
rationale is that the communication style of the words that accompany a visualization can influence how much effort a learner puts into trying to see how the visualization relates to the words. The first row of Table 22.4 shows there is support for the personalization principle across 15 experimental tests carried out in conjunction with our lab, including learning about lightning in a computer-based multimedia presentation (Moreno & Mayer, 2000b, Experiments 1 and 2), the human respiratory system in a computer-based multimedia presentation (Mayer, Fennell, Farmer, & Campbell, 2004, Experiments 1, 2, and 3), botany in a multimedia game (Moreno & Mayer, 2000b, Experiments 3, 4, and 5; Moreno & Mayer, 2004, Experiments 1a and 1b), engineering in a multimedia game (Wang, Johnson, Mayer, Rizzo, Shaw, & Collins, 2008, Experiment 1), and chemistry in an online tutoring system (McLaren, DeLeeuw, & Mayer, 2011a, Experiments 1a and 1b; McLaren, DeLeeuw, & Mayer, 2011b, Experiments 1a and 1b). Overall, students who received multimedia lessons with conversational style performed better on subsequent transfer tests than did students who received lessons with formal style, yielding a median effect size of $d = 1.0$. An important boundary condition is that the effects of personalization were strong for lower knowledge learners but not higher knowledge learners (McLaren, DeLeeuw, & Mayer, 2011a, 2011b; Wang et al., 2008).

The voice principle is that people learn better from multimedia lessons when the words are spoken by a friendly human voice rather than by a machine voice. The rationale is that the quality of the instructor’s voice can influence how much effort a learner puts into trying to see how the verbal explanation meshes with the visualizations. For example, there is support for the voice principle across six experimental comparisons carried out in our lab, involving computer-based lessons on lightning (Mayer, Sobko, & Mautone, 2003, Experiment 2), mathematics word problems (Atkinson, Mayer, & Merrill, 2005, Experiments 1 and 2), and solar cells (Mayer & DaPra, 2012, Experiments 2a and 2b). Overall, students who received multimedia lessons with human voices performed better on transfer tests than did students who learned with machine voices, yielding a median effect size of $d = 0.7$, as shown in the second line of Table 22.4. A potential boundary condition is that the positive effects of voice are reduced when there are other negative cues, such as when an onscreen agent shows low levels of movement and gesture (Mayer & DaPra, 2012).

The embodiment principle is that people learn better from multimedia lessons when an onscreen agent exhibits human-like gestures and movements. The third line of Table 22.3 shows that across five experimental comparisons carried out in our lab involving a multimedia lesson on solar cells (Mayer & DaPra, 2012), there is preliminary evidence for the embodiment principle, with a median effect size of 0.6. The embodiment principle did not apply when there were negative social cues, such as machine voice, suggesting an important boundary condition.
Where Does Multimedia Instruction Work?

There are many venues in which visualizations can be added to verbal instruction, including book-based lessons (e.g., illustrated textbooks), computer-based multimedia lessons (e.g., narrated animations or interactive simulations and games), face-to-face lessons (e.g., PowerPoint presentations), and lessons with hand-held devices (e.g., illustrated textbooks on Kindle or interactive games and simulations on a cell phone). Additionally, visualizations can be incorporated into many different subject areas, including reading (e.g., graphics in phonics instruction), writing (e.g., graphic organizers for planning an essay), mathematics (e.g., concrete manipulatives to help represent word problems or computational procedures), science (e.g., interactive simulations and games for complex systems), history (e.g., figures, graphs, and maps to depict data, or photos and drawings to depict places and people), and second language learning (e.g., as photos, drawings, video, animation to aid in vocabulary learning).

Much of the research presented in this chapter focuses on illustrated text, narrated animations, and computer-based interactive games, using content mainly in science and mathematics. However, there are encouraging signs that researchers are beginning to identify how best to incorporate visualizations in a full range of venues and subject areas (Hattie, 2009).

PRACTICAL IMPLICATIONS

The major practical implications of this chapter are the 11 design principles for multimedia lessons listed in Tables 22.2, 22.3, and 22.4, and described in the previous sections. Extraneous processing during learning can be reduced by using the coherence, signaling, redundancy, spatial contiguity, or temporal contiguity principle. Essential processing during learning can be managed by applying the pretraining, segmenting, or modality principle. Generative processing during learning can be fostered by incorporating the personalization, voice, or embodiment principle.

These practical implications are examples of evidence-based practice—basing instructional methods on research evidence rather than on conventional wisdom, opinion, speculation, fads, or doctrine (Clark & Mayer, 2016). In particular, the 11 design principles summarized in this chapter constitute the fruits of a research strategy that can be called basic research on applied problems (Mayer, 2011) or use-inspired basic research (Stokes, 1997). In basic research on applied problems (or use-inspired basic research), researchers seek to accomplish two overlapping goals—to contribute to theory (such as a cognitive theory of multimedia learning) and to solve a practical problem (such as how to design effective multimedia instruction). For example, the research reported in this chapter helps develop a cognitive theory of multimedia learning by examining boundary conditions of design principles as predicted by the theory and helps improve design of multimedia instruction by providing evidence-based principles.

Stokes (1997) shows how the theoretical and practical goals of researchers yield four different research scenarios—pure basic research occurs when researchers seek to contribute to theory but not to practice, pure applied research occurs when researchers seek to contribute to practice but not to theory, poor research occurs when researchers seek to contribute neither to theory or practice, and use-inspired basic research occurs when researchers seek to contribute to both theory and practice. Stokes refers to this final scenario as Pasteur’s quadrant and shows how it can lead to many advances in
Instruction Based on Visualizations

• 497

science and practice. Instead of viewing applied and basic research as two ends of a continuum, it makes more sense to view them as complementary research goals. This is the approach taken in this chapter, and more widely within this Handbook.

FUTURE DIRECTIONS FOR RESEARCH ON MULTIMEDIA INSTRUCTION

The research presented in this chapter points to the potential value of adding visualizations to verbal instruction—an instructional approach that I have called the promise of multimedia learning (Mayer, 2009). Advances in computer and communication technologies have created renewed interest in adding visualizations to verbal instruction in order to help people learn. This renewed interest in the promise of multimedia learning can be seen as the continuation of a 350-year old quest dating back to Comenius’s well-reasoned rationale for the world’s first illustrated textbook in the mid-1600s. Based on the current state of the research on learning with visualizations, I propose some promising and unpromising directions for future research.

Some promising directions for research on visualizations include (a) the discovery of new evidence-based principles for multimedia design, (b) the determination of whether evidence-based principles of multimedia design apply in authentic learning situations and with delayed testing, (c) the pinpointing of the boundary conditions of multimedia design principles indicating when they most strongly apply, and (d) including the role of motivation and metacognition in multimedia learning based on in-process measures.

First, the technology of multimedia instruction (i.e., the development of multimedia instruction) is emerging at a faster rate that the science of multimedia instruction (i.e., evidence-based principles for multimedia design and research-based theory). As textbooks continue to migrate from paper-based to computer-based media, students increasingly are exposed to instructional multimedia games and simulations, and multimedia instruction becomes available on hand-held technologies and virtual reality technologies, there is a need for research on how best to design effective multimedia instruction. Although researchers have made substantial progress in identifying some preliminary design principles, developers need more guidance to help them design effective multimedia instruction.

Second, most of research in this chapter is based on short-term laboratory studies with immediate tests, so complementary work is needed that examines how design principles apply in schools and training venues courses with actual students.

Third, research is needed that pinpoints the boundary conditions under which design principles are most (and least) likely to be effective, including for which kinds of learners, which kinds of content, which kinds of outcome measures, and which kinds of learning contexts. The search for boundary conditions should be guided by learning theory, and can be used to help test and refine current theories of how people learn from words and pictures.

Fourth, most of the studies cited in this chapter used posttests of learning achievement. Given the increased recognition of the role of motivation and metacognition in academic learning, future research should add in-process measures (including brain-based measures, eye-tracking measures, and other physiological measures) to examine the role of cognitive, metacognitive, and motivational processes during learning.

Some unpromising directions include educational research focusing on media comparison studies and unscientific studies of the development of new visualization technologies. First, in media comparison studies, researchers compare learning with
one medium (e.g., TV, film, books, computers, virtual reality) versus another medium. Although media comparison studies have a long history in educational research (Saat-tler, 2004), the consensus among educational researchers is that our field does not need more media comparison research (Clark, 2001). Clark (2001) and Clark and Feldon (2014) note that instructional media do not cause learning, but rather instructional methods cause learning. The same medium can be used for different instructional methods; thus, useful research focuses on which instructional methods best foster learning rather than on instructional technology per se. This can explain why media comparison studies generally fail to yield strong effects (Hattie, 2009).

Second, some of the research on multimedia instruction focuses mainly on describing the development of new technologies, describing how people use a new technology, or showing that people like using a new technology. In some cases, researchers seek to show that a new technology fosters learning, but they do not include a control group, do not have appropriate learning measures, or do not have sufficient numbers of learners. In some cases, the main data are descriptions of what students do as they learn. Instead, what is needed is high-quality, scientific research that answers the research question at hand. When the research question concerns which instructional methods are most effective for producing learning with multimedia instruction, the most appropriate methodology is experiments that met the criteria of random assignment of participants, experimental control including a control group, and appropriate measures that focus on learning outcome and learning process (Mayer, 2011). Overall, there are encouraging signs for continued advances in our understanding of how to use visualizations to help people learn.

**NOTE**

1 Preparation of this chapter was supported by grant N00014162046 from the Office of Naval Research. The author’s email address is mayer@psych.ucsb.edu.

**REFERENCES**


Comenius, J. A. (1887). *Orbis pictus. [The world in pictures: Nomenclature and pictures of all of the chief things in the world and men’s employments therein.]* Syracuse, NY: C. W. Bardeen, Publisher. [Originally published in 1658.]


