Title

Learning from Instructional Video: Theoretical and Empirical Explorations

Author

Geller, Emma Harlan

Publication Date

2016

UNIVERSITY OF CALIFORNIA

Los Angeles

Learning from Instructional Video:
Theoretical and Empirical Explorations

A dissertation submitted in partial satisfaction of the
requirements for the degree Doctor of Philosophy
in Psychology

by

Emma Harlan Geller

2016

ABSTRACT OF THE DISSERTATION

Learning from Instructional Video:
Theoretical and Empirical Explorations

by

Emma Harlan Geller

Doctor of Philosophy in Psychology

University of California, Los Angeles, 2016

Professor James W. Stigler, Chair

Historically, the study of instruction and the study of cognitive processes involved in learning have proceeded in different fields with little overlap. Issues of random assignment and experimental control have made studies of cognition in educational settings difficult, expensive, and prone to failure, while well-controlled experiments often occur in contrived laboratory situations. More recent efforts have attempted to bridge these issues by using more educationally relevant materials in laboratory studies, and by bringing well-controlled experiments into the classroom. The recent proliferation of instructional video is one tool that provides an inexpensive, easily controllable, and ecologically valid way to study the effects of instruction on cognition and learning.

Much of the existing research on multimedia learning has focused on cognitive processes and instructional principles to improve learning, but there is little attention paid to the semantic understanding that students develop when they read or watch a lesson. The text comprehension literature offers a model for how understanding develops, and may be fruitful for studying video instruction. In three studies, I investigate the usefulness of text comprehension as a model for understanding instruction.

In the first study, I assess the effect of adding information that increases the local coherence of a lesson. These experiments show that additional information improves retention for some question types, but it is not clear whether the effect is due to increased salience of the particular information rather than local coherence *per se*. In the second study, I investigate the effect of segmenting and labeling portions of a lesson. These 5 experiments fail to show consistent patterns of results, but highlight the importance of participants' prior knowledge and engagement with the lesson. In the last study, I evaluate the effects of advance organizers on learning. This experiment highlights the importance of the specific materials used and suggests that well-cited findings from the literature may not be easily replicable. Though these studies provide mixed results overall, they contribute to our understanding about how people learn from video by highlighting the importance of the materials we teach, the tests we use, and the prior knowledge of the participants.

The dissertation of Emma Harlan Geller is approved.

Robert A. Bjork

Keith Holyoak

Louis M. Gomez

James W. Stigler, Committee Chair

University of California, Los Angeles

2016

This dissertation is dedicated to the Francis W. Parker Charter Essential School in Devens, MA
for stoking my intellectual curiosity and for showing me what
effective teaching and learning can look like.

Table of Contents

LIST OF FIGURES

LIST OF TABLES

ACKNOWLEDGEMENTS

I'd like to thank my advisor, Jim Stigler, whose general disregard for "the rules" has greatly widened and enriched my graduate school experience. Thank you for teaching me to think like a storyteller and to tackle the good, interesting problems rather than the easy ones. Your advice and mentorship has been invaluable. Another huge thank you goes to Karen Givvin, whom I admire so much; you are the kind of teacher and mentor I want to be. You have no idea how much your support has meant to me.

Many others have had a hand in shaping my ideas and my work. To my committee members – Keith Holyoak, Bob Bjork, and Louis Gomez – thank your for your invaluable feedback and support, and for reminding me that non-significant results are an important part of research and can still provide important insights. To Gerardo Ramirez, Ji Son, Belinda Thompson, and everyone who ever attended a TALL meeting, thank you for helping me through all my half-baked ideas and for making lab meeting the thing I looked forward to every week. To Bob and Elizabeth Bjork and the CogFog crew, thank you for your consistent and constructive feedback (you saw more finished work than my own lab did) and for feeding my body and mind every Friday morning. To my research assistants who helped enormously with creating materials and collecting and analyzing data: Binh Nguyen, Haley McNamara, Fiona Davies, Lucy Cui, Jared Schwartz, and Georgia Potter. I could not have done this without your help. Thank you also to the friends who commiserated and celebrated with me at every milestone along the way: Jared, Ines, Janine, Lauren, Courtney, Jeff, Mikey, Melissa, Natsuki, Genna, and especially KP. We made it!

I also owe a huge debt of gratitude to the many teachers and mentors in my life who have helped me develop my intellectual curiosity and who have demonstrated what effective teaching

and learning can look like. To all of my teachers at Parker, but especially Matt Smith, Dawn Crane, Heidi Kulik, Janice Gallagher, Deb Merriam, Jon Bohannon, and Diane Kruse, thank you for guiding me through the most formative educational experience of my life. I still sing the praises of Parker to anyone who will listen, but it was really you who created that experience for me and showed me how to view learning as a process of reflection and self-improvement. I still strive to be half the teachers you were for me. To my mentors at GW, Jerry Feldman, Larry Medsker, Stephen Dopkins, and especially Cornelius Bennhold, thank you for introducing me to research and showing me how I could combine my many interests in the psychological study of learning in math and science classrooms. Had it not been for Physics 001, I might never have ended up in graduate school! All of you have deeply shaped the way I think about teaching and learning and the role research can play in improving education.

A special thank you also goes to my husband, Mike, who has been with me every step of the way. Our first date was just days before the start of my first year, and neither of us knew then that we'd be here together (and married!) five and half years later. I'm excited to be moving on with you by my side.

Lastly, thank you to my parents for encouraging me to go to graduate school at all. I was ready to settle when you told me to aim high; I'm so glad you did.

VITA

EDUCATION

Ph.D., The University of California, Los Angeles, expected June 2016
Psychology
Committee: James W. Stigler (chair), Keith J. Holyoak, Robert A. Bjork, Louis M. Gomez
Dissertation: *Learning from Instructional Video: Theoretical and Empirical Explorations*

M.A., The University of California, Los Angeles, March 2012
Psychology
Committee: Philip J. Kellman (chair), James W. Stigler
Thesis: *Practice with High-Level Cognitive Tasks Produces Changes in Basic Information Extraction*

B.A., Summa Cum Laude, The George Washington University, May 2010
Psychology
Honors Thesis: *Attitude, Structure, and Problem-Solving: An Analysis and Review of Two Teaching Methods for Introductory Physics*
Advisor: Cornelius Bennhold

AWARDS AND ACHIEVEMENTS

2012-2014	Advanced Quantitative Methods in Education Research, Two-Year Fellow Funded by IES Grant R305B080016 to the University of California, Los Angeles ($60,000 in stipends, travel funds, fee remissions)
2012-2013	UCLA Graduate Research Mentorship ($23,000)
2011 & 2012	UCLA Graduate Summer Research Fellowship ($4,600 & $6,000)
2010-2011	UCLA University Fellowship ($20,000)
2009	Phi Beta Kappa member
2008	GWU Luther Rice Fellowship ($5000)

PUBLICATIONS

Geller, E. H., Son, J. Y., & Stigler, J. W. (in preparation) Conceptual explanations and understanding fraction comparisons. *Mathematical Thinking and Learning.*

Stigler, J. W., Geller, E. H., & Givvin, K. B. (2015) Zaption: A platform to support teaching, and learning about teaching, with video. *Journal of e-Learning and Knowledge Society, 11*(2): 13-25.

Bufford, C. A., Mettler, E., Geller, E. H., & Kellman, P. J. (2014) The psychophysics of algebra expertise: Mathematics perceptual learning interventions produce durable encoding changes. *Proceedings of the* 36th *annual meeting of the Cognitive Sciences Society* (CogSci); 2014 July 23 – 26; Quebec City, Canada.

CONFERENCE PRESENTATIONS

Geller, E. H. & Stigler, J. W. (2015, July). Using advance organizers to improve learning from video. Poster presented at the 37th Annual Meeting of the Cognitive Science Society (CogSci); July 23-25; Pasadena, CA.

Geller, E. H., Son, J. Y., & Stigler, J. W. (2014, May). Explanations and understanding of fractions. In J. W. Stigler (Chair), *Relational Understanding in Mathematics: Issues and Interventions*. Symposium conducted at the 26th Annual Convention of the American Psychological Society (APS), San Francisco.

Bufford, C. A., Mettler, E., Geller, E. H., & Kellman, P. J. (2014, May). Capturing mathematics perceptual learning through psychophysics. Poster presented at: 26th Annual Convention of the American Psychological Society (APS); May 22-25; San Francisco, CA.

Geller, E. H., Son, J. Y., & Stigler, J. W. (July, 2013). Explanation and fractions: How preference for types of explanations affects learning. Poster presented at the 35th Annual Meeting of the Cognitive Science Society (CogSci); July 31 – August 3; Berlin, Germany.

INVITED ADDRESSES

Geller, E. H. (May 2015) Using Cognitive Psychology to Improve Multimedia Instruction. UCLA Library *Talks to Educate the Staff* series.

Geller, E. H. (March 2012) Perceptual Learning & Applications to Education. Psychology in Action outreach event at Palisades Charter Elementary; Pacific Palisades, CA.

TEACHING EXPERIENCE

Courses Taught as Primary Instructor
- Introduction to Psychology – Summer 2014
- Freshman Seminar with Writing credit – Spring 2015
  - Title: *Hack Your Brain: Using Cognitive Psychology to Become a Better Learner*

UCLA Graduate Teaching Assistant
- Psychological Statistics – Fall 2011
- Research Methods in Psychology – Spring 2012
- Introduction to Cognitive Psychology – Winter 2012 and Fall 2013
- Cognitive Laboratory – Winter 2014
- Mind Over Matter: History, Philosophy and Science of the Brain (general education department) – Fall 2014, Winter 2015, Fall 2015, Winter 2016

Guest Lectures
- Cognitive and Linguistic Development – Winter 2013 and Winter 2014
  - Lecture title: *Expertise and Learning*
- Introduction to Cognitive Psychology – Fall 2013
  - Lecture topic: *Problem Solving*

Chapter 1: Introduction

For most of the history of education, research on the principles that guide instruction and the study of mental processes involved in learning has largely proceeded in two separate fields (Glaser, 1991; Mayer, 2005). Cognitive psychology has spent many decades building a science of learning in which carefully controlled experiments elucidate the cognitive processes students engage in when they attempt to learn information (Glaser & Bassok, 1989). These studies are strong tests of theories about the science of learning, but they often take place in contrived laboratory situations in order to maintain experimental control (Mayer, 2008). On the other hand, the field of education research has spent a great deal of energy studying natural classroom situations and evaluating the relationship between instructional methods and students' achievement outcomes. This field of study is more ecologically valid for understanding the impact of instruction, but has been at best correlational and offers little explanation about what kinds of thinking and processing students engage in as a result of instruction (Mayer, 2005).

One reason there has been little overlap between the science of learning and the science of instruction is that it is very difficult to ensure experimental control over classroom teaching. When new instructional techniques are to be tested in a classroom, it can be very difficult to get instructors to change the way they teach; they may be uncomfortable or unfamiliar with the intended technique, and the implementation of the instructional method may vary substantially across teachers. Additionally, because students are nested within classrooms, randomized controlled trials require enormous numbers of students to draw statistical conclusions. These issues have historically made experimental work in educational settings expensive, difficult, and especially prone to failure.

However, more recent efforts have attempted to bridge this gap by using educationally relevant materials in laboratory studies and by bring well-controlled experiments into the classroom. In line with these developments, the recent proliferation of instructional video provides an opportunity to study instruction in a systematic and inexpensive way. Video is relevant to the study of instruction not simply because it is a new technology, but because video lessons are repeatable and controllable. Much like classroom instruction, video lessons are intended to inform and support students in learning and understanding material relevant to classroom achievement. But unlike face-to-face classroom instruction, the same lesson can be played repeatedly, in exactly the same way for hundreds of thousands of students. What's more, students can be randomly assigned to instructional conditions at the individual – rather than the classroom – level. This repeatability is what makes video instruction the perfect laboratory for studying instruction.

1.1 Cognitive Theory of Multimedia Learning

The first place one might look to understand learning from video is Richard Mayer's (2005) Cognitive Theory of Multimedia Learning (CTML). CTML is a framework for understanding how students learn from materials that contain both pictures and words. It is based in both the science of learning and the science of instruction. The theory states that learning is a result of three cognitive processes: selecting, organizing, and integrating information. To select information, we must attend to the appropriate stimuli. We then organize the selected information by building coherent cognitive structures, like hierarchical trees or propositional networks. Lastly, we integrate the organized information into our existing knowledge by making appropriate connections between structures and prior knowledge.

CTML also makes three assumptions about the processing of information in a lesson (Mayer, 2005). The first is the *dual-channels* assumption: the idea that working memory uses two distinct channels – visual and verbal – to process incoming information. Verbal information (i.e. ideas communicated through words) may be either auditory (spoken) or visual (printed text), while pictorial information (i.e. ideas communicated through pictures or animations) is only visual. The second assumption is that our working memory has a *limited capacity*; we cannot hold every bit of information at once, so we make decisions about what to rehearse and what to ignore as we process incoming information. Lastly, people learn through *active processing* of material rather than passive reception. In other words, learning is not merely receiving and repeating information, but instead actively organizing and integrating it.

With these assumptions in mind, Mayer (2005) argues that the purpose of instruction is to manage the amount and kinds of cognitive processing students engage in. He identifies three kinds of cognitive processing that are related to cognitive load theory (Sweller et al., 1998; Sweller, 2005): extraneous, essential, and generative. Extraneous processing is mental effort that is unnecessary because it does not support the instructional goal of the lesson. This could include reading information that is interesting but not relevant to the instructional goals, or mapping correspondences between figures and captions that are on different parts of a page or screen. Essential processing is the mental effort that is attributed to the complexity of the task but that is needed to accurately represent the material. Generative processing is the mental effort associated with *making sense* of the information by organizing it and integrating it with prior knowledge.

Given our cognitive limitations, the goals of good instructional design is to reduce extraneous processing, manage essential processing, and foster generative processing. If these goals are not achieved, then learners experience cognitive overload and key ideas or inferences are likely to be lost. Mayer (2005) suggests ten principles of multimedia learning that can be used to design lessons that minimize the "bad" processing and maximize the "good" kind, such as removing unnecessary or redundant information (Harp & Mayer, 1997; Mayer & Jackson, 2005; Moreno & Mayer, 2000), highlighting or signaling important words and ideas (Harp & Mayer, 1998; Mautone & Mayer, 2001), and allowing lessons to be self-paced (Mayer & Chandler, 2001; Mayer, Dow, & Mayer, 2003).

Overall, the cognitive theory of multimedia learning provides a theoretical basis for investigating *methods* of instruction. The assumptions of the theory – dual processing channels, limited capacity, and active processing – help guide decisions about what kinds of instructional moves are likely to support student learning (particularly transfer). Mayer's (2005) meta-analysis of instructional methods identifies those methods that have strong supporting evidence from cognitive psychology regarding their effectiveness at promoting transfer.

What the theory does not provide, however, is a model of how students build a *semantic* representation of the lesson. For example, CTML suggests that removing seductive details from a lesson can improve students' learning, but it does not offer an explanation for how students build appropriate representations of the content of the lesson or make connections with prior knowledge. In other words, CTML addresses the processing and capacity limits of our cognition, but does not provide a model of *what students do* when they read or listen to an expository text. The literature on text comprehension proves fruitful in addressing this question and will be described in the following section.

1.2 Text Comprehension

CTML makes strong claims about what kinds of instructional decisions affect the types of processing in which students engage, but it does not offer much explanation about the objects and products of those processes: how do students select what details to process? How is information organized and integrated with prior knowledge? Researchers studying text comprehension developed computer models that could, in fact, extract the meaning of a text on the basis of a limited number of comprehension principles (eg. Kintsch & van Dijk, 1978). These models were somewhat sensitive to the processing capacity of humans, but they made different assumptions about cognitive processing because they were based primarily on semantic representations of text alone.

Video lessons contain more information than text by virtue of being presented in multiple modalities, but expository text and instructional video do share some important similarities. First, both unfold over time and in a specified, sequential order. Traditional studies of learning and memory have relied on the use of lists as stimuli, and though the items on a list must necessarily be studied in some order, that order is not generally important to the meaning of the list (aside from interleaving effects). In contrast, expository texts and instructional lessons are presented in a specific order to maintain the coherence of ideas. There are multiple ways to organize a text or lesson, but those ways are limited if the lesson is to make sense.

A second important similarity is that expository texts and video lessons are designed *for the purpose of instruction.* Certainly one may potentially learn from any kind of text or video, but not all texts or videos are designed to be educational. The purpose of an expository text or lesson, however, is to inform the audience of some information. Models of the way this information is presented and processed in text form may help us to understand how students make sense of video lessons when the content is auditory or pictorial. There may be important differences between the way text and video are processed, but text comprehension is a useful jumping-off point and source of comparison.

Third, both texts and lessons are organized around a conceptual structure. Unlike list learning, where individual items may not be related in any superordinate way, ideas in a text or lesson are supposed to be organized in a coherent structure, often hierarchically. Although propositions appear in a sequential format when reading or listening to a lesson, their interrelationships generally have some global structure that organizes the domain. One goal of the student should be to see or discover this organization in order to make sense of what he is seeing or hearing. When students are able to represent the organization of a text or lesson, we would say that they have comprehended the text.

Kintsch's (1988) integration-construction model of text comprehension provides a useful starting point for exploring comprehension of expository materials. In this model, Kintsch (1986; 1988) proposes that readers contend with 3 levels of analysis in any text: the surface structure, the text base, and the situation model. The surface structure is the physical stimulus with which participants interact. The text base is the semantic representation of the text and represents the result of the construction phase of construction-integration. The integration phase results in what Kintsch calls the situation model, which consists of the connections between the text base and prior knowledge. The richer a reader's background knowledge, the more opportunities there are to build connections between the text base and existing knowledge. These connections help to both organize the text base and suggest relevant inferences beyond the text base.

In short, comprehension of the text base depends primarily on the quality and structure of the text itself, whereas comprehension of the situation model depends primarily on the reader's prior knowledge; both kinds of comprehension can influence each other (Kintsch & van Dijk, 1978). Importantly however, only a well-developed situation model is associated with deep understanding of a text (Kintsch, 1994); someone with a strong text base could accurately recall the text, but would not necessarily *understand* it well if that text base is not well integrated into an elaborated situation model. An important point about these two kinds of understanding is that two readers may have equally elaborate understandings of the text base, but very different understandings of the situation model if their background knowledge is not the same.

1.3 Coherence

Coherence is paramount to text comprehension, and it may be defined in both a local and a global sense (Graessar, Singer, & Trabasso, 1994). Local coherence refers to 'referential overlap' across propositions in a text. In other words, the degree of local coherence reflects whether adjacent clauses or sentences are about the same thing(s). If a sentence were to switch referents in the middle, it would not longer be coherent. Likewise, if the referents were to switch mid-paragraph, we would lose a sense of what this part of the text is *about*. This level of textual analysis is what Kintsch and van Dijk (1978) refer to as the *microstructure* of discourse. Local coherence can be improved by making referents (especially pronouns) explicit and filling in inferential or explanatory gaps in the writing (Britton & Gulgoz, 1991; McNamara, Kintsch, Songer, & Kintsch, 1996).

Global coherence, on the other hand, refers to the 'topic of discourse' and its organization; Kintsch and van Dijk (1978) call this the *macrostructure* of discourse. The macrostructure of a text organizes and preserves the hierarchy of information in a text by reducing unnecessary or redundant ideas, generalizing from specific to general propositions, and constructing facts from local knowledge chunks. For example, in a text about different kinds of heart disease, many details about different kinds of heart disease are presented in a locally coherent way, but the macrostructure contains few specific details and instead identifies organizing ideas, such as the distinction between congenital and acquired disease, or between diseases caused by blockages and those caused by holes (Kintsch, 1994). The global coherence of a text can be improved by signaling the macrostructure in the text itself (using headings, paragraph breaks, and bolded words, for example) and by changing or improving the learner's prior knowledge on the topic (McNamara et al., 1996). When learners have high prior knowledge, there are many candidate ideas by which to organize and connect the new information, whereas learners with low prior knowledge are more dependent on only the information contained in the text.

One key component of Kintsch and van Dijk's (1978) processing model of comprehension is that readers must process information in cycles of small chunks and with limited working memory. A reader can only hold a small part of the text in working memory at any given time, and while it is held there they must determine whether the propositions are both locally and globally coherent. Local coherence can be determined within a chunk by assessing whether the propositions contain overlapping referents, but global coherence is determined over cycles of processing where the coherence of later chunks is assessed in relation to previous cycles and to prior knowledge of the domain (Kintsch, 1994). For example, a reader will assess whether that text is globally coherent (i.e. has a theme or topic) if later sentences are related to earlier ones. In order to make this judgment, she must have *some* knowledge of the domain to know whether the ideas are related or not or else rely entirely on the surface structure and general text-processing strategies.

The literature on text comprehension identifies several ways of manipulating the coherence of a text with differing effects on different measures of learning. I will discuss some of these manipulations below along with related ideas from education and cognitive psychology.

I have divided these ideas into two broad classes: those that affect coherence and learning when presented *before* instruction, and those that take place *during* instruction.

1.4. Effects on Coherence Prior to Instruction

The studies and ideas discussed below differ in the extent to which they explicitly provide and organize information, but they all concern *preparation* for learning and describe ways that learning effects can be different even when the lesson students experience is the same. The literatures discussed below are prior knowledge effects, advance organizers, and preparation for future learning.

1.4.1 Prior knowledge

Prior knowledge can affect both the local and global coherence of a text, as well as the situation model that readers develop. For example, local coherence can be affected when authors use multiple words for the same concept, but the reader does not know that those words are referring to the same thing (Britton & Gulgoz, 1991). Global coherence can be affected when authors omit explicit connective inferences and readers do not have enough background knowledge to generate those inferences (Kintsch, 1988; Kintsch, 1994). The absence of explicit organizing statements can mean that readers fail to develop a coherent text base, or their situation model is poorly integrated with the text base. When readers have high prior knowledge, they have many candidate ideas to organize what they are reading, and their general knowledge about the domain helps them make inferences that are not included in the text itself (Graesser, Singer, & Trabasso, 1994). Readers with low prior knowledge must rely more heavily on the text to make sense of what it is about and may not be able to generate certain inferences that are not made explicit in the text itself.

In Mayer’s (2005) review of principles of multimedia learning, he identifies *pre-training* as a principle of instruction that is useful for managing essential processing of the lesson. This principle suggests that providing learners with additional knowledge prior to instruction improves their retention and transfer of the lesson. Most of these studies involve lessons on functional systems such as how brakes work or how lightning forms, and the pre-training helps to familiarize subjects with the names, locations, and characteristics of the components of the system. When subjects are familiar with this information before the lesson, they outperform subjects who did not receive pre-training (Mayer, Mathais, & Wetzell, 2002; Mayer, Mautone, & Prothero, 2002). Pre-training improves learning because subjects can dedicate their processing resources to building a cause-effect model of the system, rather than learning the relevant terminology *and* trying to understand the system.

While prior knowledge may improve learning from a lesson or text, McNamara et al. (1996; also McNamara, 2001) have found that readers with different levels of background knowledge benefit from different degrees of coherence in the text itself. They measured readers’ free recall of the text and accuracy on text-based questions, as well as their ability to make key inferences, solve problems related to the text, and sort key words. When readers’ background knowledge was low, they benefitted from a coherent text, but when background knowledge was high, readers learned more from a minimally coherent text. Thus the learning outcomes for a given text are dependent on both the coherence of that text as well as the prior knowledge of the reader.

1.4.2 Advance organizers

Advance organizers are relevant, subsuming concepts that are presented before a lesson or text in order to facilitate the retention of meaningful text (Ausubel, 1960). Specifically, advance organizers provide a conceptual framework that is at a higher level of abstraction than the text itself, which helps students organize the information as they read. Not all material provided before a text or lesson is considered an advance organizer; common prefaces to textbook chapters frequently give historical or chronological information about the material to be learned, but this does not function to guide the structure of what students are about to learn (Ausubel, 1960). Although there has been some controversy over the effectiveness of advance organizers in improving meaningful learning, extensive reviews of the literature have concluded that advance organizers are effective when they are used and measured appropriately (Ausubel, 1978; Mayer, 1979).

Advance organizers might function in two ways: to enhance the global coherence of the text base, or to enhance the reader's prior knowledge of the domain (Mannes & Kintsch, 1987). When organizers are consistent with the organization of the text but provide headings or labels for superordinate ideas, they may simply affect the global coherence of the text by providing a framework by which to organize and store the ideas. Rather than asking students to identify the global structure as they read, an advance organizer can reduce the processing demands by providing organizational structures before reading.

On the other hand, some advance organizers may *not* enhance the global coherence of the text but still improve readers' retention and transfer of material by enhancing their prior knowledge structures. Mannes and Kintsch (1987) tested such a theory by manipulating whether the information in an outline was organized in a way that was consistent or inconsistent with the text. In this study, participants were to read an article about the use of bacteria in industry, but first studied an outline that contained additional information about bacteria. In the consistent condition, the headings of the outline matched those of the article they would later read, but in the inconsistent condition, the headings matched an encyclopedia entry on bacteria. Identical information was contained in both outlines; only the organization was different.

Mannes and Kintsch (1987) found that the consistent structure lead to better cued-recall and recognition of the text, and participants' summaries more closely matched the text order, with fewer intrusions from the outline. This suggests that the consistent outline helped participants see the global coherence of the article and store more details from the article. However, the inconsistent outline lead to better performance on inference verification tasks and difficult problem-solving tests. The researchers suggest that the inconsistent outline may not have helped students identify the global structure of the article, but instead helped participants elaborate their situation model by providing advance knowledge of related ideas.

1.5 Effects on Coherence During instruction

Whereas effects on coherence prior to instruction primarily target subjects' background knowledge and knowledge structure, effects on coherence *during* instruction target the ease and depth of processing. Mayer's (2005) review of principles of multimedia instruction specifically identifies several instructional choices that affect learning by reducing or managing the processing load subjects experience. Two of these principles are particularly relevant to the present studies: signaling and segmenting.

The signaling principle of multimedia learning states that students learn more from multimedia instruction when key words and ideas are highlighted (Harp & Mayer, 1998; Mautone & Mayer, 2001; Mayer, 2008). Highlighting can be accomplished by providing an overview of the main points of a lesson at the beginning, by adding headings to sections of the lessons, or by stressing main ideas with vocal emphasis or enhanced text (such as bolding or italicizing). Signaling is an effective principle of instruction because it can guide students' attention to relevant ideas, thereby reducing the amount of extraneous processing they might otherwise engage in when watching a lesson where they must process all information and decide what is relevant.

The segmenting principle is meant to manage essential processing rather than reduce extraneous processing. This principle states that students learn more when the pace of the lesson is self-controlled than when the lesson is continuous without breaks (Mayer, 2008). Studies of segmenting compare multimedia lessons where the instruction is broken into short (1-2 sentence) chunks and the lesson pauses after each segment until the subject clicks “continue” with lessons that do not pause at all (Mayer & Chandler, 2001; Mayer, Dow, & Mayer, 2003). Segmenting the lesson in this way allows students to develop their understanding of each segment before moving on to the next one, potentially improving their understanding of the text base as well as providing an opportunity to search for connections in prior knowledge. When the lesson advances continuously, subjects experience a greater strain on working memory as they try to accomplish the selection, organization, and integration of new information at a quick pace.

1.5.1 Explanation

Another way to improve the coherence of a lesson is to provide explanation that supports the development of local and/or global coherence. Explanations can fill in unstated inferences in a text to increase local coherence and provide connections to prior knowledge. The impact of good explanations may not be obvious to the learner because most of us experience an illusion of explanatory depth: we have a tendency to think we understand the world more deeply than we actually do (Keil 2003; Rozenblit & Keil, 2002). This is often a result of activating folk understandings of phenomena and failing to recognize gaps in our own understanding (Keil, 2006). Providing a text with greater depth of explanation may improve learning of the text because it makes additional knowledge and inferences explicit, where absence of the explanation would simply activate the incomplete schemas that we mistake for deep knowledge.

Studies of self-assessed comprehension of texts show that subjects are poor judges of their own comprehension (Glenberg & Epstein, 1985; Glenberg, Sanocki, & Epstein, 1987), but that judgments of comprehension can be improved by providing examples and embedded questions in the text (Walczyk & Hall, 1989). Other research from Chi and colleagues (Chi et al., 1989; Chi et al., 1994; Chi, 2009) has shown that learning is improved when students are prompted to self-explain as they read a text. Self-explaining helps students monitor their own understanding and gaps in their knowledge (Chi et al., 1989), and prompts students to generate inferences to fill in missing information and integrate information with prior knowledge. Furthermore, there is some evidence that self-explanation may be more effective in multimedia (compared to single media) learning environments (Roy & Chi, 2005).

1.5.2 Depth of Processing & Active Engagement

Text coherence can be aided by providing explicit inferences and explanations, but this may have the adverse effect of reducing participants' active involvement in the lesson. There is strong evidence from psychology that learning and memory are strengthened when information is processed more actively and at a deeper level (Craik & Lockhart, 1972; Craik & Tulving, 1975), which may make learning activities more difficult in a way that is desirable (Bjork & Bjork, 2011). One explanation for the depth of processing effect is that deeper processing prompts subjects to make more semantic elaborations of the text (Anderson & Reder, 1979).

One way to increase depth of processing in a lesson is to change the type of task the student is engaged in during the lesson. Chi (2009) has developed a conceptual framework that differentiates the degree of active engagement in learning activities. The framework is not organized by what instructors *intend* to happen in a lesson, but rather by *what students do* (i.e. students' overt activities) and their corresponding underlying cognitive processes. In this framework, interactive activities are most effective for learning, followed by constructive activities, and lastly active activities (Chi, 2009).

Activities are active to the extent that the learner physically does something, whether searching, gesturing, paraphrasing, selecting, or otherwise engaging with the material. The cognitive processes underlying such activities include searching and activating existing knowledge, and assimilating, encoding and storing new information. Constructive activities are more elaborative than active ones because they produce outputs that go beyond the information that is given. This includes explaining, justifying, connecting, reflecting, and planning or predicting (among others). The cognitive processes that support constructive activities are inferring new knowledge, integrating new information with existing knowledge, and organizing, repairing, and restructuring knowledge. Interactive activities are *joint construction* activities where more than one person contributes to the output. These involve activities such as revising errors from feedback, arguing or defending an argument, or confronting or challenging an idea. Cognitively, interactive activities rely on processes that incorporate others' contributions into a mental model or knowledge structure (Chi, 2009).

Both multimedia learning and text comprehension are types of learning that involve active and constructive elements. Many of Mayer's (2005) principles of multimedia learning are aimed at fostering constructive activities by reducing the processing load of active activities. For example, reducing the need to search for structure by signaling important information allows the viewer to engage in inferring, integrating and organizing new information. Encouraging greater depth of processing by designing lessons to include constructive activities should therefore improve learning.

One way to make instructional activities more constructive is to intersperse questions during learning (Rothkopf, 1966; Rothkopf & Biscibos, 1967). Rickards and DiVesta (1974) demonstrated this by interpolating questions in a text and varying the frequency with which questions were posed and the depth of processing required. In their study, participants read a text passage that contained interpolated questions every 2 or 4 paragraphs, and the questions addressed either (1) facts in the text, (2) ideas in the text, or (3) organization of ideas in the text. They found that interpolated questions improved recall of the text, and that recall was especially high when the questions were frequent (after every 2 paragraphs) and addressed meaningful learning (organizing ideas in the text). Later work by Rickards and Hatcher (1977) has shown that inserted questions may serve as semantic cues for readers with low background knowledge. Rickards (1976) has also demonstrated that conceptual questions have a greater effect on recall when they are presented immediately *before* the associated text segments, and verbatim questions have a greater effect when they occur immediately *after* the text segment, but only conceptual pre-questions improved recall at a delay.

Interspersing questions in learning material seems to improve learning by increasing subjects' depth of processing and engagement with the material. Such activities, however, many not be easily completed by all subjects; one's ability to generate inferences and answer conceptual questions is strongly influenced by prior knowledge, the coherence of the text, and their interaction, as discussed above (McNamara et al., 1996). Thus one relevant question for both researchers and instructors is how to increase both explanatory coherence *and* active processing for a range of learners. Is it better to provide information that reduces the participant's need to actively construct and organize information, or is it better to increase the processing demands in a way that fosters generative processing?

1.6 Summary and Rationale

Given what we know about principles for multimedia instruction (eg. Mayer, 2005) and processes of semantic comprehension (eg. Kintsch, 1994), the goal of this dissertation is to productively combine these literatures to better understand how students develop understanding from instructional video. The experiments are organized intro three sets of studies meant to build on each other to develop an understanding of how students learn from instruction.

The first two studies assess methods of improving learning from instructional video by manipulating what is contained in the lesson itself. In these experiments, the content of the lesson was changed by separately manipulating the local and global coherence of the text base. I accomplished this by inserting text slides designed to affect the coherence of the lesson by providing greater explanation of local ideas or by signaling the global structure of the lesson. The third study was designed to assess the impact of advance organizers. In this experiment, the video lesson itself was not manipulated, but the information and activities that preceded it were varied by condition. The primary question addressed here was whether advance organizers are more effective when they are consistent or inconsistent with the structure of the lesson itself. The effectiveness of different advance organizers was assessed using measures of both retention and transfer. The results of these experiments may be used to contribute to a set of design principles for instructional video that are sensitive to the processing demands of the task and to the prior knowledge of viewers.

Chapter 2: General Methods

In this section I describe general methods relevant to all three sets of studies, including criteria for selecting video, the platform for adding interactive elements to video, and the design of pre- and post-tests. I then describe methods specific to each experiment separately in the chapters that follow.

2.1 Selecting Videos for Study

The recent proliferation of video means that the possible sources of video for these studies are many and varied. Websites like YouTube and Vimeo are meant to host all kinds of video (instructional or not), but contain dozens – if not hundreds – of channels dedicated to educational content. In order to narrow the options, there are several criteria for selection.

First, the videos should teach about a topic typical of a high school science course, such as biology, chemistry, physics, or psychology. These are not the only kinds of instructional videos found online, but the conceptual structure of these domains is ideal for studying the ways that students process local and global coherence while watching a lesson. These are also topics with which students tend to struggle and which are a large focus of education reform efforts.

Second, the video must be instructional. This criterion is meant to distinguish videos that simply demonstrate a concept from those that attempt to *teach* the concept by describing and explaining it. When instructors teach an in-person lesson, they do more than demonstrate phenomena; they also describe and explain the related key concepts in the domain. The videos selected for study should reflect what teachers might normally do as much as possible. This does not mean that the selected videos should *exactly* match what goes on in a classroom – I was not looking for video recordings of live lectures – but they should follow the general structure of a lesson, including introducing, demonstrating, and explaining one or more concepts related to a central topic.

Third, the videos should be between 5 and 15 minutes long. The purpose of limiting the length of the video was to avoid overwhelming the participants. In studies of both multimedia learning and text comprehension, participants rarely spent more than 20 minutes engaging with the learning material, so video lessons of approximately the same duration should be appropriate.

Fourth, the videos should be typical of what students find online with minimal search effort. In other words, the goal was not to identify the "perfect" video lesson on a given topic, but to select a lesson that students would be likely to find and watch if they were to look online for a study aid. The purpose of this dissertation is to understand how students learn from video and to test specific interventions that might improve their learning; it is not to identify the best lesson, but rather to identify principles that *improve* lessons. For this reason, it is most ecologically valid to select videos that students are likely to watch (whether they would be considered excellent lessons or not) and look for ways to improve what they learn rather than trying to change what they watch.

2.2 Adding Interactive Content

In order to test manipulations that occur during (rather than prior to) instruction, I used an online video platform called Zaption. Zaption is a web application that allows users to add text, images, and interactive elements to video. Videos may be hosted on sites like YouTube, Vimeo, PBS, and elsewhere, but Zaption lets users create lessons that require interaction from viewers, more closely simulating in-person instruction. Understanding the capabilities of Zaption may be useful for envisioning the implementation of the studies presented in this dissertation.

To create a lesson, a user begins by selecting one or more video clips and editing the duration of the clips to show exactly the desired part of the video. The user can then drop interactive elements into the lesson at precise time points in the video. The elements include informational slides that can display text, images, or drawings, or question slides with different response formats (open response, multiple choice, check all, numerical answer, or drawn response). Elements can be stacked at the same time point to create clusters of questions or slides that viewers see and respond to before the video continues, or they can be spread out at different time points. They can also be placed either on top of or next to the video window to control whether or not the viewer can see what is displayed in the video while they answer the question. All forced-choice question formats allow the user to determine whether viewers will get feedback on their answers. In addition, if a user wishes to create two slightly different versions of the same tour, she does not need to painstakingly create multiple tours from scratch. Instead, she can use the *clone* function to create an exact replica of the original tour and then make edits to the elements or video clips where she pleases.

These features make Zaption a powerful research tool because they offer a great deal of control over the type and timing of elements added to a lesson and because of the ability to easily create multiple conditions of the same lesson that differ in systematic ways.

2.3 Design of Pre- and Posttest Measures

Pre- and posttest measures were meant to assess prior knowledge and learning. Learning will here be defined in two ways, consistent with Mayer’s (2005) measures of multimedia learning: retention and transfer.

Pretest measures function as tests of prior knowledge. In order to minimize the effects that a pretest may have on processing a lesson, all pretest items reflected general knowledge of the lesson topic but not specific details within the lesson. This was accomplished by asking participants to rate the degree to which they understood or could describe a complex process in that domain and to indicate whether they were familiar with vocabulary specific to that domain. For example, a pretest before a lesson on the nervous system might ask participants to rate how well they can explain the way neurons communicate and to indicate (yes/no) whether they can define terms like action potential, neurotransmitter, afferent/efferent, and limbic system. The purpose of the pretest is to measure participants' self-reported understanding of a topic before the lesson, so that prior knowledge can be correlated or controlled for in learning outcomes.

Posttest measures consisted of two parts – retention items and transfer. The exact questions on a given posttest depend on the content of the lesson, but their general form can be used in many contexts. Retention tests consist of a general recall question asking participants to summarize the main points of the lesson, as well as specific cued-recall and/or recognition items that probe memory for specific details from the lesson. For example, after a lesson on the nervous system, participants can be asked to summarize the main points of the lesson by identifying the major parts of the nervous system and what they do, or to describe the process by which neurons communicate in as much detail as they can. Cued-recall and recognition items can ask more specific questions like "What part of the neuron does the 'listening' to other neurons?" or "What type of neurotransmitter increases the likelihood of an action potential in the post-synaptic neuron?"

Transfer tests consist of items that require participants to make inferences and generalizations beyond the information in the lesson itself. As described in Mayer's (2005) work, four broad types of questions can guide transfer tests: troubleshooting, redesign, prediction, and concepts. An example of a troubleshooting question about the nervous system might be

“Suppose a person touches a hot stove but doesn’t feel any pain. Why might that be?” A prediction question could ask “Is it possible for someone to experience paralysis and still feel sensation in the paralyzed area? Why or why not?” An example of a conceptual question could be “What causes a neuron to fire an action potential?” Lastly, a redesign question might ask, “What could you do to increase the probability of a neuron firing an action potential?”

The exact questions on a transfer posttest should be answerable based on the lesson (i.e. they would not require prior knowledge beyond what was presented), but the answers should not be directly presented in the lesson. The examples shown above presume that the lesson content does not provide these answers directly but that it does provide information on neuronal communication and on the difference between motor and sensory systems. Likewise, the example retention questions also presume that the lesson would specifically mention that the dendrites “listen” to other neurons and that excitatory neurotransmitters make action potentials more likely in the post-synaptic neuron.

Separating posttest performance in this way allows me to make stronger claims about the type of understanding students develop from instruction. Students’ accuracy on retention items should reflect the quality and structure of their representation of the information contained in the lesson, (the text base). Performance on transfer items, on the other hand, should reflect the integration of the text base with their prior knowledge (the situation model).

Chapter 3: Effects of Explanation on Local Coherence

This chapter investigates the effect of increasing local coherence of a video lesson on participants' ability to recall and apply key concepts in the lesson. Local coherence is the degree to which adjacent sentences and clauses are related or refer to the same thing. It is common for authors to leave local coherence gaps in their writing that can be filled by generating an inference. For example, consider these sentences: "The neuron sends an electrical impulse down the axon. When it reaches the terminal buttons, neurotransmitters are released into the synapse." There is no explicit overlap of referents in the two sentences, but a reader with a good grasp of English can easily infer that "it" in the second sentence refers to the electrical impulse.

This coherence gap is easily resolved, but not all gaps are the results of underspecified referents. A common source of difficulty for novices in a domain is discovering which vocabulary words are synonyms of one another (for example, "action potential" and "electrical impulse"). Yet more difficult are coherence gaps that require some background knowledge of the domain. Texts can be made more coherent by filling in these gaps with explicit statements of inferences and by using the same word to refer to a concept throughout the text (or clarifying which terms are interchangeable) (Kintsch & Van Dijk, 1978; Britton & Gulgoz, 1991). In other words, additional explanation of key ideas or terms should increase the local coherence of a text and lead to better recall.

In these studies of instructional video it is not possible to alter the narration or on-screen text in a lesson, but Zaption does allow us to insert definitions and explanations in places where they would increase local coherence. The purpose of the experiments below is to assess the impact of changes to local coherence on learners' retention and transfer for the lesson content.

Experiment 1

This first study was meant to test the effect of increasing local coherence by pausing the video lesson and adding text slides to clarify inferential gaps. The comparison of interest was whether the video with additional information (informative pauses) would help participants retain more of the lesson that the original video (no pauses). In order to tease apart the benefits of the additional information and the effect of simply pausing the video, a control condition included all of the pauses from the enhanced video, but without the content (non-informative pauses). A fourth condition had participants complete the posttest *before* watching the video (pretest only) to demonstrate that participants did, in fact, learn from the video lesson.

Based on effects of local coherence in text comprehension, I expected to see that participants' retention would be highest when they saw informative pauses and lowest with no pauses at all. Given that the pause itself does not add clarifying information, the non-informative pauses condition should not be different from no pauses. The pretest only condition should have the lowest performance overall because participants would not have watched the video lesson before trying to answer the questions. If pretest performance were the same as the no pauses condition that would be evidence that participants did not actually learn from the video.

Methods

Participants

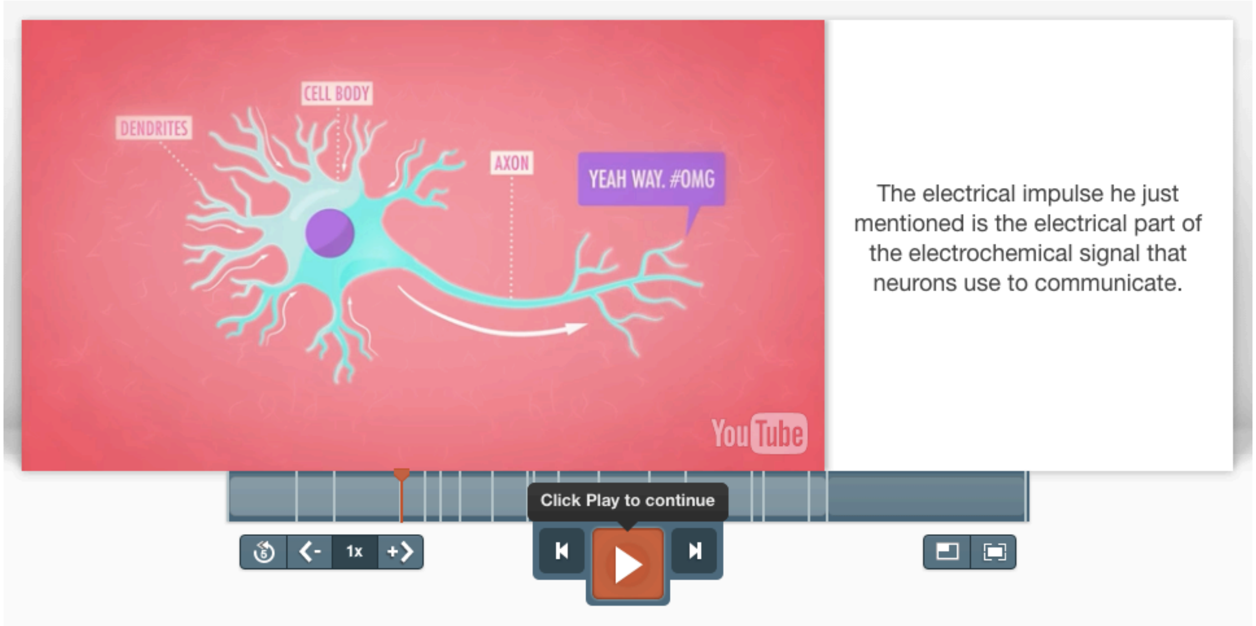
Participants were 228 workers on Amazon Mechanical Turk (134 female; M_{age} =36.5 years) who were each paid $1 for participation.

Materials

Video lesson. Participants watched a short video lesson on the role of chemical messengers in the body (original YouTube link: https://www.youtube.com/watch?v=W4N-7AlzK7s). The lesson covered several topics including the basics of neuronal communication (action potentials and neurotransmitters), the endocrine system, and how neurotransmitters and hormones differ. The video was 9 minutes and 34 seconds long, and was professionally produced for a YouTube channel called Crash Course.

Informative Pauses. Pauses were inserted into the video at points when I felt students would have trouble understanding key points of the lesson. Sometimes these trouble spots were due to unclear overlap of referents, such as using "electrical signal", "neural impulse", and "action potential" interchangeably without clarifying that they refer to the same thing. Other times, pauses were used to repeat important information that was presented quickly, such as definitions of excitatory and inhibitory neurotransmitters, which were only displayed onscreen for 4 seconds in the original video. Additionally, some slides reviewed longer segments of the lesson by restating several key points rather than simply repeating the last sentence. Lastly, some pauses corrected common misconceptions about neuronal communication, such as the belief that the synapse is part of the neuron (rather than the gap between neurons). These pauses were inserted completely at my discretion and were based on my intuitions from teaching lessons like this one in a typical face-to-face classroom. A complete transcript of the lesson, including the text in the informative pauses, can be found in Appendix A.

There were 16 total pauses in the video. On average, pauses were 35.6 seconds apart, but gaps between pauses ranged from 6 seconds to 66 seconds. In the informative pause condition, some pauses contained multiple elements "stacked" on one another. This was done to provide the intended information while keeping the text readable (not overwhelming the display window with text). Slides had an average of 22.4 words, while pauses (consisting of 1-4 slides each) contained an average of 36.4 words. An example of an informative pause is shown in Figure 1, below. All pauses appeared as text slides that popped up to the right of the video window. All text slides paused the video until the user clicked “play” to continue.



The electrical impulse he just mentioned is the electrical part of the electrochemical signal that neurons use to communicate.

Figure 1. A screenshot of an informative pause. The original video is shown on the left and the informative pause appears in the text slide on the right.

In total, the informative pauses added 583 words of explanation to the lesson, which itself contained 1817 words, thereby increasing the length of the lesson to exactly 2400 words (1.32 times the length of the original lesson). I was unable to record the amount of time participants spent reading the slides, but assuming that participants read at an average of roughly 200 words per minute, this would have added approximately 3 minutes to the length of the instructional video. In the empty pauses condition, every pause contained the words “Take a moment to think about what you just heard”. This phrase added a total of 160 words to the lesson, making the total length 1977 words (1.09 times the length of the original lesson). This would have added just under 1 minute to the length of the lesson, although it is likely that participants stopped reading the slides after the first few pauses.

Posttest. The posttest consisted of 8 questions designed to measure participants' retention of the material contained in the lesson. Two open response questions addressed the process of neuronal communication and the relationship between biological processes and psychological experiences. Two check-all-that-apply questions assessed students' knowledge of the anatomy of a neuron and the differences between neurotransmitters and hormones. Four multiple choice questions asked students to identify a neurotransmitter, identify a hormone, identify the definition of an excitatory neurotransmitter, and identify the endocrine gland that receives direct input from the central nervous system. The complete posttest can be found in Appendix B.

Different question types were scored in different ways appropriate to the type of response given. Multiple-choice questions were each worth 1 point (either correct or incorrect), check-all-that-apply questions were worth 4 and 5 points (based on the number of options correctly accepted or rejected), and open response questions were worth 2 and 3 points (based on the quality of the answer or explanation). The total score was calculated as the proportion correct where each question was weighted equally.

Design & Procedure

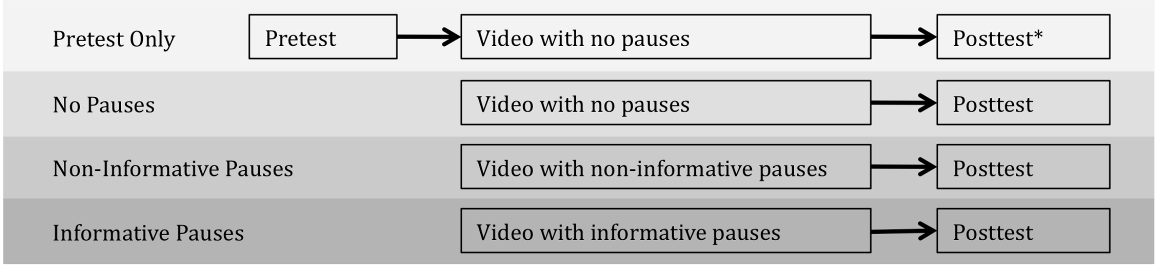


Figure 2. Procedure for Experiment 1. *The posttest from the Pretest Only condition was not included in the analyses below.

Participants were randomly assigned to one of 4 between-subjects conditions: no pauses (*n*=63), empty pauses (*n*=60), informative pauses (*n*=50), or test only (*n*=55). The general procedure is outlined in Figure 2. In the pretest condition, participants answered the 8 posttest questions *before* watching the video. Their instructions were as follows:

> “In this tour you will learn about how our brains rely on chemicals to communicate thoughts and emotions. Before you begin, we'd like to ask you some questions to see what you already know about chemicals in the body and brain.
>
> Please answer all of the questions, even if you have to guess. It's okay if you don't know the answers - you will learn about this material in the video following the questions. You must click "submit" for your answers to be recorded.”

In the no pauses, non-informative pauses, and informative pauses conditions, participants did not take a pretest, but instead began by watching the video lesson with either no pauses, non-informative pauses, or informative pauses, according to their condition. They saw the following instruction at the beginning of the video:

> “In this tour you will learn about how our brains rely on chemicals to communicate thoughts and emotions. The video moves quickly, so feel free to click Pause if you need time to stop and think. You will not be able to rewind the video as you watch, but you may pause it at any time. Press play when you are ready to begin.”

After participants watched the video lesson, they saw the following instructions:

> “Let's see how much you've learned. Your answers will help us know whether this lesson was helpful in teaching this material. Please answer all of the questions, even if you have to guess. You must click "submit" for your answers to be recorded.”

All participants then answered the 8 posttest questions. For the purposes of this study, the posttest scores from the Pretest Only condition were excluded from the analyses below.

Results

Overall, posttest scores (proportion correct) were approximately normally distributed with a mean of 0.59 and a standard deviation of 0.19. Figure 3 shows the mean proportion correct for check-all, multiple-choice, and open responses questions, as well as the total retention score, by condition.

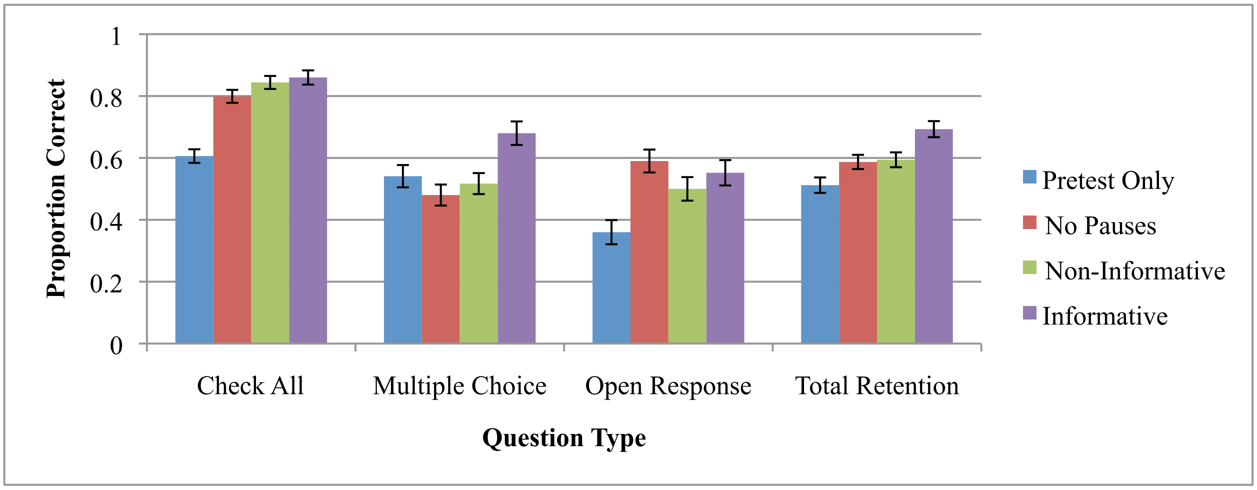


Figure 3. Mean proportion correct by condition and question types in Experiment 1. Error bars indicate standard errors.

Total retention score was analyzed using a one-way ANOVA which showed that the effect of condition on total retention score was statistically significant: $F(3,224)=8.60$, $p<0.001$. Post hoc analyses showed that the pretest only condition scored significantly lower than the informative pauses condition ($t(103)=-5.24$, $p<0.001$), but was not significantly lower than the non-informative pauses ($t(113)=-2.35$, $p=0.099$) or the no pauses condition ($t(116)=-2.23$, $p=0.158$). The overall retention score was nearly identical between the no pauses and non-informative pauses conditions ($t(121)=-0.21$, $p=1.00$). The informative pauses condition scored significantly higher that all three other conditions: non-informative pauses: ($t(108)=2.83$, $p=0.032$); no pauses: ($t(111)=3.14$, $p=0.015$): pretest only ($t(103)=5.24$, $p<0.001$). All post hoc analyses used Bonferroni corrections for multiple comparisons.

A second analysis looked at the effect of condition on performance for different question types (multiple choice, check-all, and open response) using a one-way MANOVA. The effect of condition was significant on all three question types: Check All: $F(3,224)=27.31$, $p<0.001$; Multiple Choice: $F(3,224)=5.71$, $p=0.001$; Open Response: $F(3,224)=6.81$, $p<0.001$. Post hoc analyses showed that different question types produced different patterns of performance by condition.

For Check All questions, informative pauses, non-informative pauses, and no pauses were not statistically different from each other, but all three scored significantly higher than the pretest only condition (Informative: $t(103)=7.74$, $p<0.001$; Non-informative: $t(113)=7.61$, $p<0.001$; No Pauses: $t(116)=5.84$, $p<0.001$). The lack of a difference by condition may have been due to an overall ceiling effect. Figure 4 is a histogram of Check All scores that shows this.

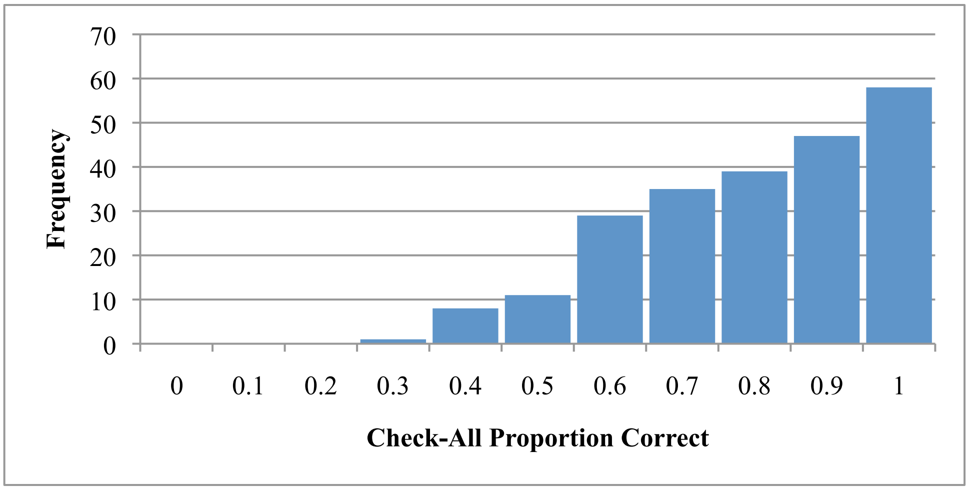


Figure 4. Histogram of Check-All scores in Experiment 1.

For Multiple Choice questions, non-informative pauses, no pauses, and pretest only were not significantly different from each other, but informative pauses scored significantly higher than all three (Non-informative: $t(108)=3.04$, $p=0.010$; No Pauses: $t(111)=4.01$, $p=0.001$; Pretest only: $t(103)=2.81$, $p=0.050$).

For open response questions, the informative pauses, non-informative pauses, and no pauses were not significantly different from each other, but all three were at least marginally higher than the pretest only condition (Informative: $t(103)=3.49$, $p=0.005$; Non-informative: $t(113)=2.45$, $p=0.063$; No Pauses: $t(116)=4.42$, $p<0.001$). All post hoc analyses used Bonferroni corrections for multiple comparisons.

Discussion

Overall, the type of pause included in the lesson only made a difference on retention for multiple-choice questions. For Check All and Open Response questions, the effect of pause was only significant because it included the pretest only group; post hoc analyses indicated that the other three conditions (informative pauses, non-informative pauses, and no pauses) did not differ. This shows that participants did actually learn something from the video, but it suggests that pausing and providing extra information did not improve their ability to answer these questions. In the case of the Check-All questions, the lack of an effect may have been due to a ceiling effect for those questions.

On the other hand, the type of pause made a huge difference in performance on multiple-choice questions. The analysis suggests that retention was not improved by watching a video with non-informative pauses, nor by watching a video at all (given that Pretest Only performance was not statistically different from No Pauses). Only the informative pauses condition significantly improved retention, which implies that the additional information – rather than the pauses, per se – was what contributed to the improved performance on those questions. Watching the video improved performance on Check-All and Open Response questions (relative to the Pretest Only condition), but only Multiple-Choice questions benefitted specifically from the inclusion of informative pauses.

This study provides some support for the idea that increasing local coherence can improve retention, but the evidence is limited and does not provide any information about the effect of coherence on transfer. It's also possible that the effects of local coherence are muted because performance on some questions is at ceiling. If retention is at a maximum immediately following the lesson but decays at different rates for different conditions, then increasing the retention interval might result in differences between conditions that I could not detect here.

Experiment 2

In Experiment 1 I found evidence that adding informative pauses to the lesson increased viewers' retention of the lesson primarily by improving their performance on multiple-choice questions. However, performance on Check-All questions showed evidence of a ceiling effect which may have obscured or muted the effect of pauses on retention overall. The results of Experiment 1 also do not indicate whether those effects persist over a longer retention interval, and do not show whether participants can apply the concepts in the lesson to novel questions.

Experiment 2 addressed these issues by including a short (3-min) distracter task as well as a transfer posttest. The purpose of the distracter test was to decrease overall performance on the posttest in order to eliminate the ceiling effect on the Check-All questions, which could reveal an effect of condition that was not detectable in Experiment 1. The posttest was also expanded to include 3 open response transfer questions. The goal of these questions was to assess how well participants could apply relevant information from the lesson to answer novel questions about the nervous system.

Additionally, I eliminated the Pretest Only condition from the first experiment. The results from Experiment 1 showed that participants did learn from the video, but the comparison of interest was between the types of pause included in the lesson. In Experiment 2, participants were randomly assigned to one of three video lesson conditions: *no pauses*, *non-informative pauses*, or *informative pauses*. These video lessons were identical to those used in Experiment 1.

Methods

Participants

Participants were 102 workers on Amazon Mechanical Turk (59 female; M_{age} =37.9 years) who were paid $1 for participation.

Materials

Video lesson. The video lesson in this study was the same as in the previous study. The lesson covered several topics including the basics of neuronal communication (action potentials and neurotransmitters), the endocrine system, and how neurotransmitters and hormones differ. The video was 9 minutes and 34 seconds long, and was professionally produced on for a YouTube channel called Crash Course. All pauses were identical to those used in the first study.

Distracter Video. The distracter video was a 3-minute clip from another Crash Course video on an unrelated topic: the agricultural revolution. Participants were asked to pay attention to the advantages and disadvantages of agriculture that were presented in the clip, and to report as many of those advantages and disadvantages as they could remember after watching the clip.

Posttest. The posttest was nearly identical to the posttest used in Experiment 1, and contained 7 retention questions and 3 transfer questions. One question from the original posttest ("What does it mean to say that "everything psychological is biological"?") was eliminated because it was difficult to agree on a scoring scheme and the question did not produce a wide enough range of responses. The other 7 questions from the original posttest (which can be found in Appendix B) were included in this experiment. The total retention score was calculated as the proportion correct where each question was weighted equally.

Three transfer questions were added to the original posttest, and were meant to assess whether participants could use information from the lesson to reason about novel questions and/or situations. The three open response transfer questions were:

- Are psychoactive drugs more like neurotransmitters or hormones? Why?

- If someone doesn't show a fearful response in a dangerous situation, why might that be? List all the reasons you can think of.
- Why does injury to the spine only cause paralysis below the injury?

These questions were scored on a scale from 0-2 or 0-3 depending on the question. Higher scores were given to answers that gave reasons or explanations that drew on material from the lesson, such as the differences between neurotransmitters and hormones, or the hypothalamus-pituitary-adrenal pathway (both of which were described directly in the lesson). Two questions were scored on a 0-3 scale and one was scored on a 0-2 scale because there was less variability in the possible answers participants could give. The overall transfer score was calculated as the proportion correct, where each question was weighted equally.

Design & Procedure

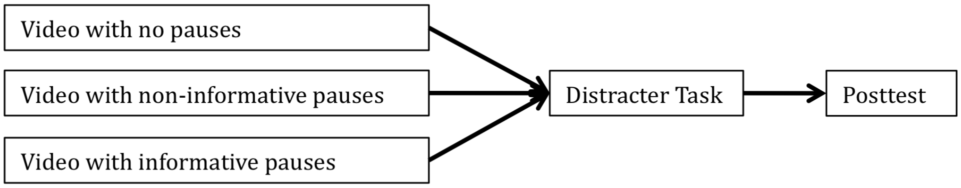


Figure 5. Design and procedure for Experiment 2.

The design of Experiment 2 is show in Figure 5. Participants were randomly assigned to one of 3 between-subjects conditions: no pauses (*n*=36), non-informative pauses (*n*=38), or informative pauses (*n*=28). The uneven group size was due to the assignment mechanism. A "randomizer" URL assigned participants to a condition, but potential participants could have clicked on the randomizer link, been assigned, and then decided not to participate before opening the lesson. Amazon Mechanical Turk does not record the number of workers who open a link but do not complete task, so it is not possible to know how many workers may have chosen not to participate after clicking on the randomizer link but before answering any questions. Participant data was only collected if participants answered at least one of the questions in the video lesson, and only complete posttests were included in the subsequent analyses. The rates of attrition (after having answered at least one question) were not different by condition (roughly 10%).

Once assigned to condition, each participant saw the same instructions:

> "In this tour you will learn about how our brains rely on chemicals to communicate thoughts and emotions. The video moves quickly, so feel free to click Pause if you need time to stop and think. You will not be able to rewind the video as you watch, but you may pause it at any time. Press play when you are ready to begin."

After reading the instructions, participants watched the video lesson, which contained informative pauses, non-informative pauses, or no pauses, according to the condition assigned. Immediately following the lesson, all participants watched a short distracter video describing advantages and disadvantages of the agricultural revolution. After watching the clip, participants were asked to list as many advantages and disadvantages of agriculture as they could. Then participants complete the posttest. Their instructions for the posttest were as follows:

> "You will now take a short test on the first video you watched. Your answers will help us know whether this lesson was helpful in teaching this material. Please answer all of the questions, even if you have to guess. You must click "submit" for your answers to be recorded."

After completing the questions on the posttest, participants were asked to rate how much they liked the lesson, how much they felt they learned from the lesson, and whether they would recommend the lesson to someone else. Participants were also given an opportunity to suggest changes to the lesson for future students.

Results

Overall retention and transfer scores non-normally distributed. Retention scores had an overall mean of 0.54 (*SD*=0.22), but showed evidence of a bimodal distribution with peaks at approximately 0.40 and 0.65 (skewness 0.273 (*SE*=0.239), kurtosis= -0.763 (*SE*=0.474)). Transfer scores were more evenly distributed across the range from 0 to 1, but were skewed to the left (*M*=0.34, *SD*=0.26, skewness=0.650 (*SE*=0.239), kurtosis= -0.198 (*SE*=0.474)). Means and standard errors for retention, transfer, and overall posttest score by condition are shown in Figure 6.

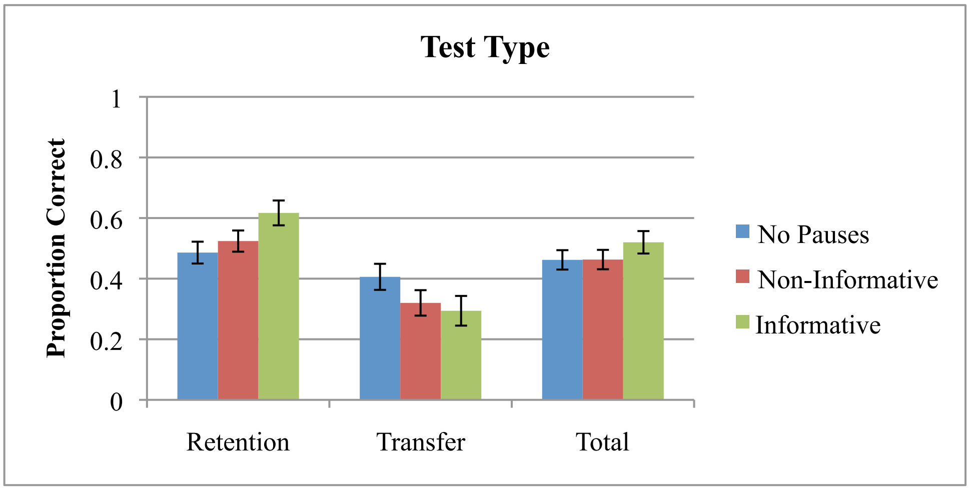


Figure 6. Mean proportion correct for retention, transfer, and total performance by condition in Experiment 2. Error bars indicate standard errors.

Overall score was analyzed using a one-way ANOVA, which showed that the effect of condition on total score was non-significant (*F*(2,99)=0.890, *p*=0.414). The retention and transfer scores were analyzed using a one-way MANOVA. There was no effect of condition on transfer (*F*(2,99)=1.74, *p*=0.181), but the effect on retention was marginally significant (*F*(2,99)=3.05, *p*=0.052). The informative pause condition (*M*=0.617, *SE*=0.041) scored marginally higher than no pauses (*M*=0.486, *SE*=0.036, *t*(62)=2.56, *p*=0.051), but was not different from non-informative pauses (*M*=0.524, *SE*=0.035, *t*(64)=1.62, *p*=0.255). Post hoc analyses used Bonferroni corrections for multiple planned comparisons.

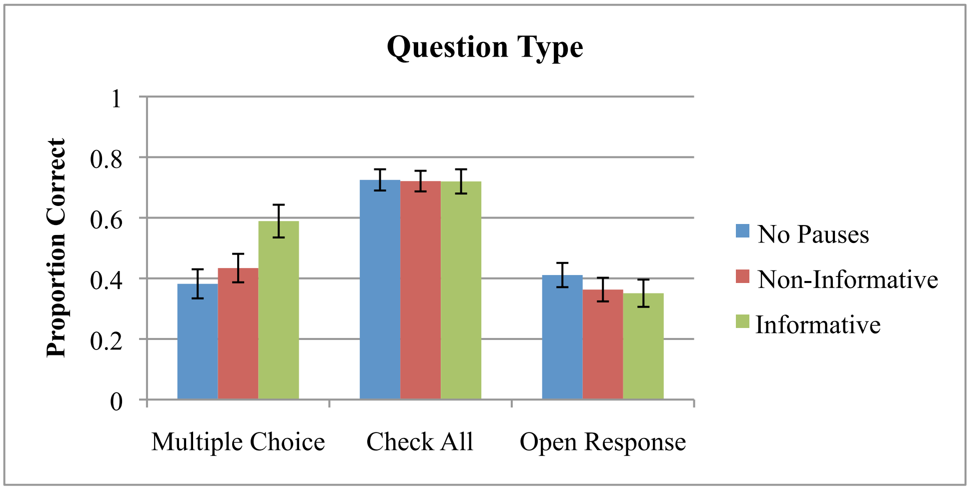


Figure 7. Mean proportion correct for multiple-choice, check-all, and open response questions by condition in Experiment 2. Error bars indicate standard errors.

Another set of analyses focused on the effect of condition on question type (multiple choice, check all, open response), as in Experiment 1. Means and standard errors for check-all, multiple-choice, and open response questions by condition are shown in Figure 7. A one-way MANOVA showed that the effect of condition was not significant for check all ($F(2,99)=0.004$, $p=0.996$) or open response questions ($F(2,99)=0.587$, $p=0.558$), but there was a significant effect of condition on multiple choice questions ($F(2,99)=4.31$, $p=0.016$). The informative pauses condition ($M=0.589$, $SE=0.054$) scored significantly higher than no pauses ($M=0.382$, $SE=0.048$, $t(62)=3.11$, $p=0.015$), but was only marginally higher than non-informative pauses ($M=0.434$, $SE=0.047$, $t(64)=2.04$, $p=0.098$). Post hoc analyses used Bonferroni corrections for multiple planned comparisons.

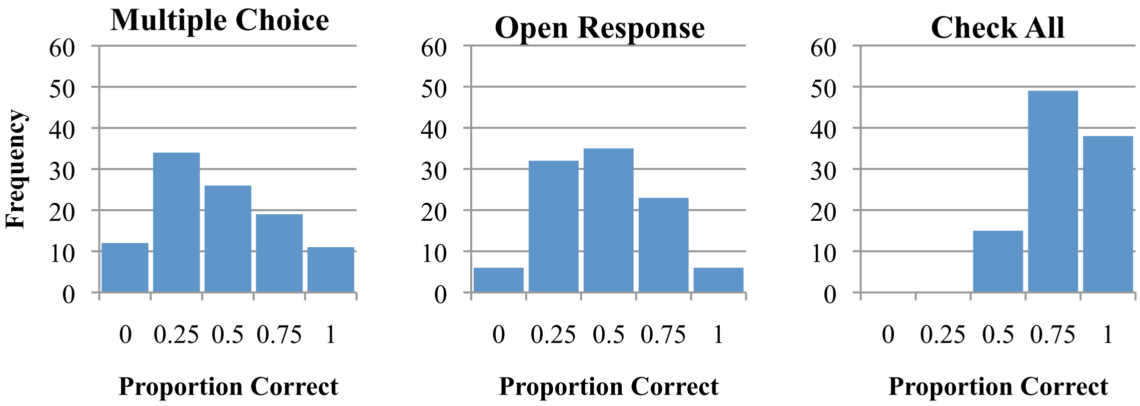


Figure 8. Histograms of proportion correct for multiple-choice, open response, and check all questions in Experiment 2. The y-axis indicates frequency.

Although overall performance was lower in Experiment 2 than in Experiment 1, scores on check-all questions remained somewhat high, which suggests that the ceiling effect may not have been eliminated. Histograms of proportion correct for each question type (displayed in Figure 8) show that multiple-choice and open responses score were relatively normally distributed, but check-all scores still show a possible ceiling effect. The peak of the distribution is lower that it was in Experiment 1, but the scores for check-all are still skewed left. This may have contributed to a muted overall effect of pause type on retention.

Participants' ratings of how much the liked the lesson, how much they felt they learned from the lesson, and whether they would recommend the lesson to a friend did not differ by condition (Like: *F*(2,99)=2.30, *p*=0.106; Learn: *F*(2,99)=0.901, *p*=0.410; Recommend: *F*(2,99)=3.03, *p*=0.053).

Discussion

As in Experiment 1, there was a small effect of pauses on retention (where informative pauses lead to higher scores than non-informative pauses or no pauses), but no effect on transfer. The results of Experiment 2 also support the finding that informative pauses increased retention for multiple-choice questions, but had no effect on check-all or open response questions. Although overall performance was lower in Experiment 2 than in Experiment 1, I was unable to eliminate the ceiling effect on check-all questions. Thus it is unclear whether the effect of pauses is obscured or whether there is no effect at all.

The effect of informative pauses on retention could have been a result of the fact that many informative pauses repeated and/or highlighted information from the lesson that was critical for answering the posttest questions. Beyond simply increasing local coherence, the informative pauses likely drew attention to that information, highlighting it as important. In the non-informative pauses condition, participants may have had a sense that something important had been presented, but without knowing exactly what information to attend to the pause was not useful for rehearsing, reviewing, or elaborating on the information presented. Further evidence to support this interpretation is the fact that the non-informative and no pauses conditions did not differ on nearly any measure. This implies that the additional information – not the pause itself – was what improved performance, although it is not clear whether this effect was due to increased coherence or increased salience of the information.

Another source of variability in the data likely came from the variety of question types. The check-all and open response questions also tested information that was highlighted in the informative pauses, but the format of those question types made it much easier to get partial credit even if one were guessing. For example, simply checking all of the boxes on the check-all questions would have earned a score of 0.70 on that question type. On the multiple-choice questions, however, it is much harder to score well by guessing; this makes them a more sensitive measure of retention than the check-all questions.

The retention test contained 4 multiple-choice questions, but the transfer test had none (only 3 open response questions), which explains why condition had an effect on retention but not on transfer. A quick analysis of open response questions also showed that the subjects who got the informative pauses scored highest on the one retention question, but lowest on the three transfer questions. This could mean that additional information might benefit retention of specific ideas from the lesson, but actually inhibit the ability to apply that knowledge to novel questions. A reverse pattern was seen with the transfer questions; although the effect was not statistically significant, participants who got non-informative pauses or no pauses scored numerically better than those who go informative pauses. Perhaps participants in the no pauses and non-informative conditions were better able to develop a gist understanding of the lesson and thus were better able to apply that thinking on the transfer questions.

General Discussion

Performance in Experiment 2 was somewhat lower than in Experiment 1, but the pattern of performance on different question types was the same. The distracter task decreased performance overall, but did not affect the pattern of responses: performance on check-all questions was lower than in Experiment 1, but there was still no effect of condition. In the same vein, participants scored lower overall on multiple-choice questions in Experiment 2 than in Experiment 1, but the pattern (a benefit of informative pauses over non-informative pauses or no pauses) remained the same. The only change in pattern was that the effect of condition on overall retention was only marginally significant in Experiment 2; this was likely due to the smaller sample sizes (roughly 55 per condition in Experiment 1, but 35 in Experiment 2).

Overall, these findings suggest that added information improves performance on multiple-choice questions, but not on check all or open response questions. These results are not altogether surprising; one might expect additional information to improve recall, especially when that information clarifies or highlights key concepts in the lesson. What remains unclear is whether this is a result of the question format, the specific information being tested, or the increased salience of that information. Interpreting these results if further complicated by the fact that performance on all question types was non-normally distributed and showed evidence of skewness, as well as possible ceiling effects. In order to draw stronger conclusions about the effects of local coherence, more work is needed to create a posttest that captures the range of participants' learning with appropriate detail and difficulty.

Limitations

This preliminary study of local coherence effects in video learning suffered from several limitations. Two of the most obvious were the length of the posttest and the sensitivity of the questions on it. The posttests in Experiments 1 and 2 contained only 8 and 10 questions, respectively, which is arguably too short of a test to capture all that participants may have learned from the lesson. Although the open response question "How do neurons communicate?" was meant to measure overall understanding of the lesson, it is difficult to assess whether participants' short and vague answers were the result of lack of retention or lack of willingness to write a more thorough response. Check-all questions required less effort to answer, but also posed problems by allowing participants to score relatively high even if they were guessing or using some other strategy. Only multiple-choice questions were resistant to these problems by being both quick to answer and relatively difficult to guess.

Another important limitation of the study was the retention interval between watching the video and taking the posttest. It is increasingly common for quizzes to be given after video lessons (especially in online courses), but one would hope that learning from video persists longer than a few seconds between watching a video and taking a test. Although not all participants in this study were performing at ceiling on the immediate posttest, we might expect different changes in memory representation as the retention interval is increased. For example, gist memory for the lesson may increase over longer intervals, which could change both the number of details participants can report, as well as perhaps their ability to apply the gist of the lesson to a novel question or situation. By testing only at a very short retention interval, the findings of this study are not easily applicable to real life situations in which several days or weeks may pass between instruction and test. Nonetheless, it provides a starting point for examining the effect of local coherence on learning.

Finally, one other potential limitation was the use of Mechanical Turk workers as participants in this study. Unlike the typical psych study participant (i.e. undergraduate students), Mturkers vary largely in age and level of education. In some ways, this may not be a limitation at all; Mturkers may be a more representation sample of the population than undergraduates. However, Mturkers have also generally been out of school for a while and it is difficult to know how focused they are on the study and how seriously they take the instructions. While Mturkers may represent the “every man” on the Internet, they may not very closely approximate student behavior. Further work is needed to know whether Mturkers and undergraduates differ appreciably in the way the engage with video lessons.

Future Directions

Future iterations of this study should focus on teasing apart the potential confounds of these two experiments, namely, the possibility that enhanced retention is due to increased local coherence or due to the increased salience of relevant information. To do this, I will also need to test the question of salience by comparing recall of information contained in the informative pauses with information from the video that was NOT highlighted in the pauses. This will require some changes to the posttest so that not all questions are aligned with information addressed in the pauses, and so that questions on the posttest cannot be easily guessed without recalling or understanding the lesson content.

Another important problem to address is the way coherence is determined. In these two experiments I relied heavily on my own intuitions about what ideas students would have trouble connecting and remembering, but I did not use a specific measure to quantify the coherence of the lesson before and after adding the informative pauses. Such measures and tools exist for text comprehension (such as McNamara's program coh-metrix: http://cohmetrix.com/) and could be used to more precisely measure the changes in coherence when information is added or removed from video narration. It is possible that coherence of narration differs in particular ways from coherence of text (in the speed or presentation or the importance of repetition, for example), but even an approximate measure could help us better understand coherence in video lessons.

In general, the goal of this work is to help students learn more from video instruction. To do this effectively, it is important to understand how student characteristics affect learning beyond the design of the lesson. Future work can take this into account by differentiating learners on the basis of prior familiarity with the material or other characteristics that may interact with pedagogical choices to produce better or worse learning. There is some evidence from studies of text comprehension that students with low prior knowledge benefit from a maximally coherent text, while students with high prior knowledge actually learn more from a *less* coherent lesson (McNamara et al., 1996; McNamara, 2001). As we try to understand the effects of local coherence in video instruction, it is important to take these kinds of individual differences into account in order to understand who benefits from coherence and when.

Chapter 4: Effects of Headings on Global Coherence

Chapter 3 focused on the effects of local coherence in a video lesson and found that providing additional information to fill in inferential gaps increased participants' retention of key ideas in the lesson. This chapter focuses on global coherence and investigates whether providing headings for segments of a lesson can improve participants' recall and/or transfer of important information in the lesson. The concept of global coherence is drawn from the text comprehension literature, and it refers to the overall organization of the text and its topic of discourse (Kintsch & Van Dijk, 1978; Graessar, Singer, & Trabasso, 1994). While local coherence focuses on the microsturcture of the text (the relatedness of adjacent clauses and sentences), global coherence focuses on the macrostructure: is it clear what the text is about? Is it clear how the different parts of the text are related to the overall point?

Studies of text comprehension have provided evidence that one can increase global coherence in a text by adding paragraph breaks, headings, and bolded words to by make the structure of the text more salient, and these changes can affect what students learn (McNamara et al., 1996). It is not the case, however, that all students always benefit from an increase in coherence; there is some evidence that students with high prior knowledge in a domain actually learn more from *less* coherent texts because the added processing difficulty causes them to engage more deeply with the material (McNamara et al., 1996). Students with low prior knowledge, however, benefit from maximally coherent materials because they are less capable of completing inferential and organizational gaps on their own.

Work from Mayer's (2005) cognitive theory of multimedia learning (CTML) also supports the use of segments and headings to improve learning. In studies of the "signaling" principle of CTML, participants' learning was increased when signals (visual or vocal emphasis) were used to reduce extraneous processing (Mayer, 2008). Some of these signals include things like providing an overview sentence at the beginning of the lesson, adding headings to sections of the lesson, and stressing key ideas vocally (Mayer, 2008). On a variety of materials (including both paper- and computer-based lessons) students learned more from lessons with these signals than lessons without them (Harp & Mayer, 1998; Mautone & Mayer, 2001; Stull & Mayer, 2007). The general conclusion of these studies is that learning is improved when important information is highlighted.

Additional work by Mayer shows the benefits of the "segmenting" principle of CTML, namely that students learn more when lessons are presented in self-paced segments than when the lesson advances automatically (Mayer, 2008). In one particular study, Mayer broke a lesson on lightning formation into 16 segments, and participants learned more when they could choose when to advance to the next segment than when the lesson advanced automatically (Mayer & Chandler, 2001). The rationale of this principle is that segmenting a lesson allows students to manage the cognitive processing that is essential for understanding the lesson (Mayer, 2008). Rather than reducing extraneous processing, segmenting helps participants manage the working memory demands associating with selecting, organizing, and integrating key information with prior knowledge (Mayer, 2005; Mayer, 2008).

In the following set of experiments, I investigate the use of both segmenting and signaling to enhance the global coherence of a lesson on kidney function. I broke a short video lesson into 12 segments and manipulated the kind of signal (heading) viewers saw before each segment. The primary manipulation of the headings involved the degree to which the headings signaled key information in each segment; some headings provided no relevant information, some highlighted key words, and others provided targeted questions. Later experiments in this chapter investigate the effects of segmenting by comparing lessons that are both segmented and signaled with lessons that are only segmented and with those that are not segmented at all.

Experiment 3

One of Mayer's (2005) principles of multimedia instruction is that comprehension is improved when important ideas and organizing concepts in a lesson are highlighted. This first experiment attempts to test this claim in an instructional video by inserting section headings that either do or do not signal the semantic structure of the lesson.

The design of the lesson draws on two of Mayer's (2005; 2008) principles of multimedia instruction: segmenting and signaling. I took at 10-minute lesson on kidney function and broke it into 12 segments that were roughly 1 minute long each, though the length varied from 23 seconds to nearly 3 minutes. Each segment of the lesson was preceded by a text slide containing a "heading" that provided some information about the lesson segment. *Number* headings simply numbered the segments from 1-12, while *Title* headings provided 2-3 word titles based on the key vocabulary in the segment, and *Question* headings posed questions that would be answered in the upcoming segment. An additional control condition did not contain any kind of heading, but paused the video and asked participants to click play when they were ready to continue.

Mayer's signaling principle suggests that students learn more from multimedia lessons where key information is highlighted or "signaled" through the use of visual or semantic cues (Mayer, 2008). This study tests the application of that principle to video lessons by examining the degree of signaling provided by different types of headings. *Number* headings do not provide much useful information to viewers because they only number the segments of the lesson without providing any semantic content that might help viewers anticipate or connect to the content of the segment. *Title* headings provide slightly more signaling utility by highlighting the relevant vocabulary for each segment, but did not guide students' attention any further; a participant might gather that the upcoming segment would describe the proximal convoluted tubule (for example), but would not know what specific information to focus on. *Question* headings, on the other hand, were written to specifically guide students' attention to the target information in each lesson segment. In fact, the questions asked in the segment headings were nearly identical to the questions on the posttest; attending to those headings should theoretically prepare students to focus on exactly the information needed to do well on the final posttest.

Based on the degree of signaling for each type of heading, I predicted that participants in the *questions* condition would score highest on the posttest, followed by the *titles* condition, and the lowest scores would be for the *numbers* condition. I did not predict that the *numbers* condition would be significantly better than the *pauses* condition (which did not provide any heading) because the *numbers* headings did not provide any useful semantic information. If the *numbers* and *pauses* conditions lead to the same performance, this would further strengthen the case that semantic signaling is important for improving students' retention of the lesson.

Methods

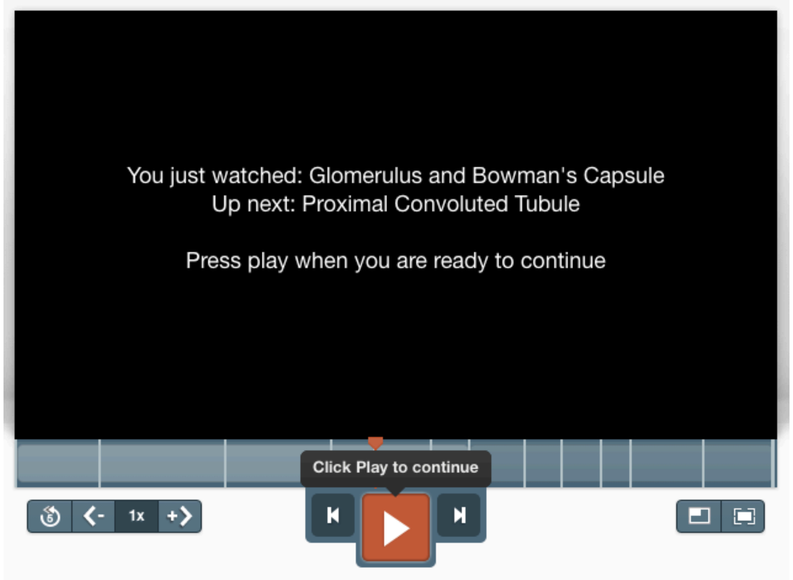
Participants

Participants were 82 workers on Amazon Mechanical Turk (46 female; M_{age} =36.4 years) who were each paid $1 for their participation.

Materials

Video lesson. Participants watched a short video lesson on the excretory system. The lesson covered several topics including the importance of osmoregulation in homeostasis, the structure of the kidney (focusing on filtering units called nephrons), and the process by which the kidneys filter blood (including each part of the nephron, what substances are reabsorbed in each part and how). The video was 12 minutes and 20 seconds long, and was professionally produced for a YouTube channel called Crash Course. I cut out a segment of the lesson that was approximately 1 minute and 30 seconds because it was not directly relevant to the overall goal of the lesson (to teach how the kidney filters blood). The final playtime of the video that participants watched was 10 minutes and 10 seconds. The original YouTube video can be found here: https://www.youtube.com/watch?v=WtrYotjYvtU

Segment headings. The video was divided into 12 instructional segments. The researcher identified segments by looking for mini topics within the lesson, such as the anatomy of the excretory system and separate segments of the nephron. The video was paused at the beginning of each segment, at which point a text slide appeared showing one of four types of headings (described below). The average length of a segment was 50.8 seconds, but segments ranged in length from 24 to 103 seconds. A complete transcript of the lesson (with segments marked) can be found in Appendix C.



You just watched: Glomerulus and Bowman's Capsule
Up next: Proximal Convoluted Tubule

Press play when you are ready to continue

Figure 9. A screenshot of a heading slide from the Titles condition.

Heading types were manipulated between subjects, so each participant saw only one type of heading throughout the lesson. In the *number*, *title*, and *question* conditions, the headings all followed the same format: “You just watched:______ Up next: ______ Press play when you are ready to continue.” An example heading slide from the *title* condition is show in Figure 9. The types of headings used in the *number*, *title*, and *question* conditions are shown in Table 1 below. In the pause condition, participants did not see any headings, and every text slide contained the same text: “Press Play when you are ready to continue.”

	Heading		
Segment	*Number*	*Title*	*Question*
1	Segment 1	Homeostasis & Osmoregulation	How does the kidney help maintain homeostastis and aid in osmoregulation?
2	Segment 2	Urea & Uric Acid	What is the difference between urea and uric acid?
3	Segment 3	Kidneys & Nephrons	What does the kidney do? How do nephrons relate to the kidney?
4	Segment 4	Glomerulus & Bowman’s Capsule	What is the main function of the glomerulus and Bowman’s capsule?
5	Segment 5	Proximal Convoluted Tubule	What substances are reabsorbed in the proximal convoluted tubule?
6	Segment 6	Renal Cortex & Medullla	What is the difference between the renal cortex and the renal medulla?
7	Segment 7	Loop of Henle	What does the loop of Henle do?
8	Segment 8	Descending Limb	What substances are reabsorbed in the descending limb?
9	Segment 9	Ascending Limb	What substances are reabsorbed in the ascending limb?
10	Segment 10	Distal Convoluted Tubule	What substances are reabsorbed in the distal convoluted tubule?
11	Segment 11	Collecting Duct	What is the function of the collecting duct?
12	Segment 12	Ureters, Bladder, & Urethra	How do the ureters, bladder, and urethra work together to excrete urine?

Table 1. Headings used in the numbers, titles, and questions conditions of Experiment 3.

Test of Prior Knowledge. In order to assess what they may have known about kidney function before watching the video lesson, participants completed a brief test of prior knowledge. Before watching the video, participants indicated how many of 10 relevant terms they felt confident they could define. These terms are shown in Table 2, below. Due to an oversight in formatting, participants had to indicate they knew at least 2 terms, but could check up to 10.

Before watching the video	*After watching the video*
Which of the following terms do you feel confident you could define?	How much did you know about this topic before watching the video?
○ Homeostasis	5 – a lot
○ Osmosis	4
○ Nephron	3
○ Glomerulus	2
○ Bowman’s capsule	1 – nothing
○ Loop of Henle	
○ Renal cortex	
○ Renal medulla	
○ Urea	
○ Urethra	

Table 2. Test of prior knowledge used in Experiments 3-7.

After watching the lesson and taking the posttest, participants were asked to rate how much of the information contained in the lesson they knew before watching the video. This question was asked after the lesson to ensure that participants knew exactly what information was contained in the lesson. If I had asked, “How much do you know about kidney function?” before watching the lesson, there would be no way to know what “a little” or “a lot” meant across participants, but when asked after the lesson participants had all been exposed to the same information and could make a more accurate judgment. Participants indicated their level of prior knowledge on a scale of 1-5, where 1 indicated they knew none of the information beforehand, and 5 indicated they knew all of the information beforehand. The average rating across all participants was 2.07 (SD=1.10).

A composite prior knowledge score was created by summing the number of terms participants indicated they could define with their self-rating of prior knowledge. Scores ranged from 3 to 14 out of a possible 15 points; the mean composite score was 4.9 (SD=2.4) and the median was 4.

Retention Test. The complete posttest (including both retention and transfer) can be found in Appendix D. The retention test was designed to measure what participants remembered from the lesson. The test consisted of 2 open responses questions (worth 5 points each) and 6 multiple-choice questions (worth 1 point each). The purpose of the open response questions was to gauge how much participants could recall freely, while the multiple-choice questions served as a sort of recognition test. The total score on the retention test was the proportion of total points correct, where each question was weighted equally.

The first open response question asked participants to describe how the kidney filters blood, using as much detail as possible. A word bank was provided to help participants remember and use key vocabulary from the lesson. Answers were scored on a scale from 1-5 according to the detail and accuracy of the answers. The average score on across all participants was 2.9 (SD=1.3). The second open response question asked participants to name and describe the parts of the nephron. An unlabeled picture of a nephron was provided as a visual aid. Answers were scored on a scale from 1-5 according to the number of parts accurately named and described. The average score across all participants was 2.1 (SD=1.6).

The multiple-choice questions primarily tested participants' knowledge of what substances were absorbed in what parts of the nephron, as well as the processes (osmosis, active transport) that were used in each stage of reabsorption. Feedback was given for each multiple-choice question (correct/incorrect) and the correct answer was also presented. Across all participants, the average number of questions answered correctly (out of 6) was 2.9 (SD=1.5).

Transfer Test. The transfer test was designed to measure whether participants could apply information from the lesson to a novel question or situation. The test consisted of 4 multiple-choice questions (worth 1 point each) and 3 open response questions (worth 5 points each). The total score on the transfer test was the proportion of points correct, where each question was weighted equally.

The first open response question asked participants to describe why kidney failure is dangerous to one's health. The second question asked how excess glucose in the blood (a symptom of diabetes) would affect the kidney's ability to filter blood and maintain homeostasis. The third question asked whether water, a sports drink, or a soda would rehydrate someone more quickly, and why. All three open response transfer questions required participants to integrate information about the filtration process. Answers were scored on a scale from 1-5 according to the validity of the answers and the amount of supporting detail provided. The average scores across all participants were 2.9 (SD=1.3) on the first question, 2.3 (SD=1.6) on the second question, and 2.8 (SD=1.4) on the third question.

The multiple-choice questions required participants to make predictions about various scenarios in which some aspect of kidney function had been altered. For example: "What would happen to the amount of water and ions reabsorbed by the nephron if someone did not have a loop of Henle?" Answers to these questions were scored as either correct or incorrect. Across all participants, the average number of transfer questions answer correctly was 1.7 out of 4 (SD=1.1).

Design & Procedure

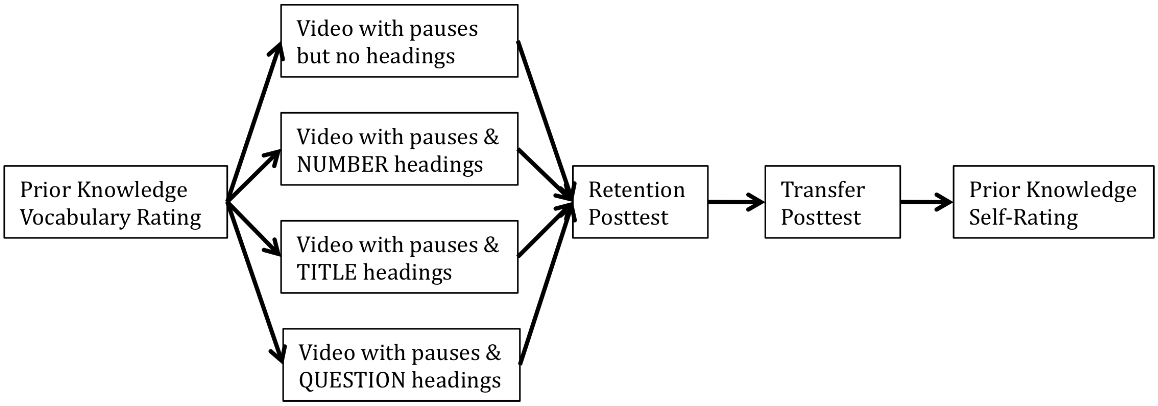


Figure 10. Procedure for Experiment 3.

The design and procedure are depicted in Figure 10. Participants were randomly assigned to one of 4 between-subjects conditions: pauses (*n*=21), numbers (*n*=19), titles (*n*=23), or questions (*n*=19). Each participant saw the same instructions:

> "In this video you will learn about the excretory (urinary) system. At the end of the video, you will take a short test to help us evaluate the quality of the lesson. Before you begin, we will ask a few questions to gauge how much you already know about this topic."

After reading the instructions, completed the vocabulary rating part of the prior knowledge test (described above). They were then shown the following instructions:

> "Please treat this video like something you would watch for fun or to learn something new. This is not a test of your aptitude or knowledge; we are just interested in how much people naturally remember from videos like this.
>
> In order to help you process the lesson, we have broken it down into shorter segments with pauses in between them. When you reach the end of a segment, you may wait for as long as you like before clicking "play" to continue."

These instructions were followed by the first segment heading (as described above and listed in Table 1), and then participants watched the video lesson. At the end of the lesson, participants were instructed that they would take a short test on the information presented:

> "Thanks for watching! You will now take a short test on the information in the video and then rate the quality of the video lesson. There will be 10 multiple choice questions and 5 open response.
>
> First, you will answer 2 open response questions about the video you just watched."

Participants saw instructions between each part of the test, indicating what type of questions they would be answering, as well as how many questions to expect. On both parts of the transfer test (multiple-choice and open response), participants were encouraged to pick the best answer or to guess if they did not know the answer.

After completing the transfer test, participants were asked to rate various aspects of the lesson including how much they enjoyed watching the video, how much they thought they learned from the lesson, whether they would recommend the lesson to someone else, and whether they would be interested in learning more about this topic. At this point, participants were also asked to rate how much of the information in the lesson they knew before watching the video; this rating was included as part of the prior knowledge score, as described above. Before finishing the study, participants also reported their age, gender, and highest level of education.

Results

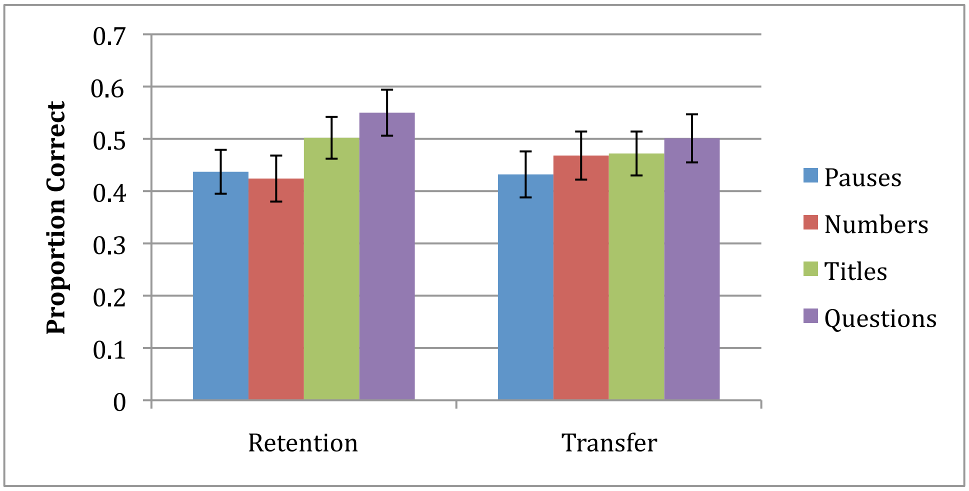


Figure 11. Proportion correct on retention and transfer posttests by condition in Experiment 3. Bars show estimated marginal means with prior knowledge as a covariate (evaluated at PK total = 4.94) and error bars indicate standard errors.

Retention and transfer scores were approximately normally distributed with an overall mean of 0.49 and a standard deviation of 0.21 for retention and a mean of 0.48 and a standard deviation of 0.20 for transfer. The effect of heading was analyzed using a one-way (heading:

pauses, number, title, question) MANCOVA with retention and transfer as separate outcomes and prior knowledge score as a covariate. The estimated marginal means for each condition (using prior knowledge score as a covariate) are shown in Figure 11. The effect of heading was not significant for retention ($F(3,77)=1.41$, $p=0.246$) or for transfer ($F(3,77)=0.408$, $p=0.748$).

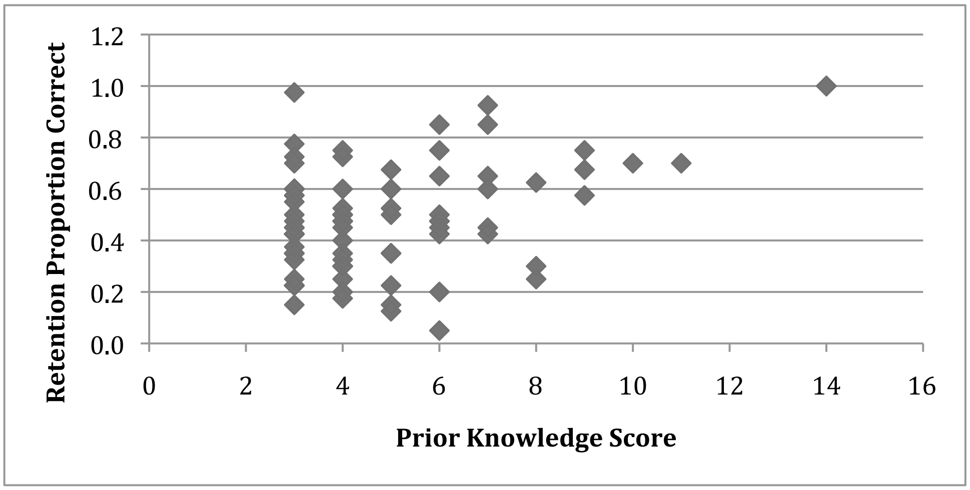


Figure 12. Scatterplot of prior knowledge score and proportion correct on the retention test in Experiment 3.

There was a significant effect of prior knowledge on retention ($F(1,77)=16.77$, $p<0.001$), but no effect on transfer ($F(1,77)=1.05$, $p=0.309$). Post hoc analysis showed that the correlation between prior knowledge and retention was moderately positive and statistically significant: $r(82)=0.415$, $p<0.001$. A scatterplot of prior knowledge score and proportion correct on retention is shown in Figure 12. There were two high outliers who scored 14 on the prior knowledge measure and 1.00 on the retention test; removing them from the correlation analysis reduced the estimate of Pearson's r, but the relationship was still significant: $r(80)=0.251$, $p=0.025$.

Discussion

Although the effects of heading on retention and transfer were not statistically significant, numerically we see the pattern that was expected: *questions* tended to outscore *titles*, which tended to outscore *numbers*. Also, *numbers* and *pauses* were not different on retention (though *pauses* scored lower than *numbers* on transfer). Also as expected, prior knowledge was significantly related to retention scores (though surprisingly not related to transfer). One possible reason that the effect of heading was not significant is simply because there is a lot of variability across subjects in terms of prior knowledge and overall score. A within-subjects design might help us detect an effect of heading where the variation in prior knowledge is better controlled.

Experiment 4

The results of Experiment 3 showed the expected pattern of question headings leading to greater retention than title headings, and title headings showed greater retention than number or pause headings. However the effect was not statistically significant, and may have been dampened by the variability in prior knowledge across subjects. Experiment 4 addresses this possibility by using a within-subjects design. In this experiment, all participants saw all 4 heading types (pause, number, title, question) an equal number of times across the 12 lesson segments (each heading type preceded 3 out of the 12 total segments).

In Mayer's (2008) work, signals are generally effective because they direct participants' attention. In the between-subjects design in Experiment 3, the different types of headings may have directed participants' attention to different degrees, but the level of attention overall was probably similar for all segments of the lesson. In a within-subjects design, however, the contrast of different heading types might mean that participants' attention to the lesson should differ from segment to segment. In other words, the within-subjects design might have a greater effect on the distribution of attentional resources over the course of the lesson, which could make differences between heading types more obvious.

For example, when a participant in Experiment 3 saw *question* headings for all 12 segments, there was no reason to pay more attention to some segments than others. However, if *question* headings direct attention more efficiently than *title* or *number* headings, then participants who see only 3 *question* headings in the within-subjects design should attend more efficiently to those segments than to segments preceded by other heading types. Specifically, we would predict that *questions* and *titles* would draw more attention than *numbers* or *pauses* because they direct viewers to look for specific terms or ideas in the lesson. Also, as predicted in Experiment 3, *questions* should direct attention more efficiently than *titles* because they direct students to look for a specific bit of information in the segment.

To test this prediction, I revised the posttest to include questions specific to each segment of the lesson so that I could compare performance on questions from segments preceded by different types of headings. All participants saw all heading types, and the order of heading types was counterbalanced across participants. I predicted that performance would be highest for *question* segments and lowest for *number* segments, and that there would be no difference between *number* and *pause* segments.

Methods

Participants

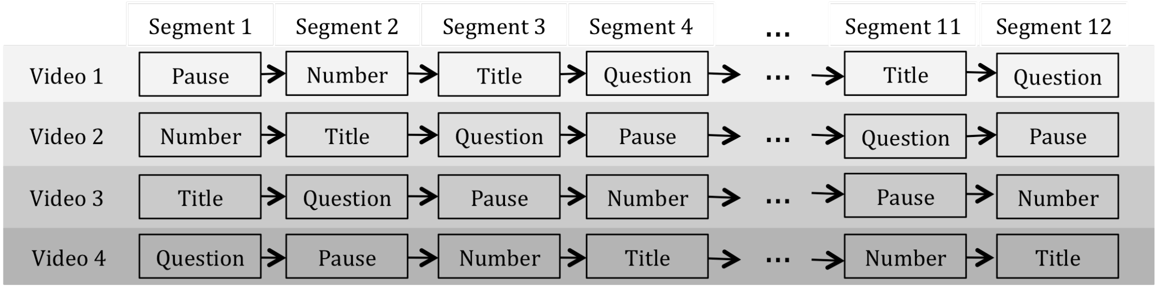
Participants were 80 workers on Amazon Mechanical Turk (53 female; M_{age} =37.2 years) who were paid $1 each for participation.

Materials

Video lesson. I used the same video lesson as in Experiment 3. The video was 10 minutes and 10 seconds long, and it covered information about homeostasis, osmoregulation, and kidney function. The lesson was divided into the same 12 segments as in Experiment 3.

Segment headings. The headings used in this study were the same as those used in Experiment 3 (see Table 1 for a complete list), but each participant saw 3 of each heading type, rather than all

12 of the same type. I created four counterbalanced lessons in which each segment was preceded by each type of heading across all conditions. The order of heading types was not counterbalanced; rather than using a Latin Square, all headings proceeded in the same order: pause, number, title, question. Each of the four lessons simply began at a different point in that sequence and cycled through it 3 times (to get 12 total headings). The pattern of headings across segments for each of the four videos can be found in Figure 13 below.



	Segment 1	Segment 2	Segment 3	Segment 4	…	Segment 11	Segment 12
Video 1	Pause	Number	Title	Question	…	Title	Question
Video 2	Number	Title	Question	Pause	…	Question	Pause
Video 3	Title	Question	Pause	Number	…	Pause	Number
Video 4	Question	Pause	Number	Title	…	Number	Title

Figure 13. Schedule of headings for each segment in each of four counterbalanced video lessons used in Experiment 4.

The heading slides were further altered to separate the “You just watched: ___” and “Up next: ___” parts of the text on separate slides. In Experiment 3, both parts were included on the same slide to highlight the phrasing of the headings and to help participants keep track of their place in the lesson. In this study, however, I wanted to reduce opportunities for participants to compare headings across segments. Had I included different heading types on the same slide, participants would have been able to directly compare headings of adjacent segments, and this would have made heading differences more salient. Although they likely remembered at least some of the headings from previous segments, separating the text was intended to reduce the probability of direct comparisons across segments.

Test of Prior Knowledge. The prior knowledge test contained the same ten terms and self-rating question as in Experiment 3, but the formatting of the vocabulary sections was modified to include “none of the above” as an option. This meant that participants could indicate their ability to define 0-10 terms, and the possible composite scores ranged from 1 to 15 points. Actual scores ranged from 1 to 14 with a mean of 2.96 (SD=1.6) and a median of 3.

Retention test. The complete posttest, including both retention and transfer, can be found in Appendix E. The retention test was changed to include only multiple-choice questions in order to make the within-subjects comparison. The open response questions from Experiment 3 required participants to report and/or integrate information from the entire lesson. Given that participants in this study saw all headings types over the course of the whole lesson, different heading types would not likely affect such global retention questions across segments. Instead, multiple-choice questions were developed to target specific segments of the lesson, allowing me to compare segments preceded by one heading type with those preceded by another heading type.

There were 16 questions total and each question corresponded to a specific segment of the lesson (4 segments had 2 questions each, and the other 8 segments had one question each). By making questions specific to certain segments of the lesson, it was possible to create a retention score for each heading type by taking the proportion of questions corresponding to each heading type that were answered correctly. For example, 3 of the 12 segments in every lesson were preceded by *question* headings: the 4 questions corresponding to those 3 segments make up the *question* retention score – likewise for the 4 questions corresponding to *title* segments, the 4 questions corresponding to *number* segments, and the 4 questions corresponding *pause* segments.

All questions on the retention test were answered directly in the lesson. None of the questions required participants to make inferences or predictions about the information presented; they simply needed to remember the information given in the lesson.

Transfer test. The transfer test contained only the multiple-choice questions from Experiment 3. Open response questions were eliminated from this test for the same reason they were eliminated from the retention test. The four multiple-choice questions happened to correspond to a single heading type in all four of the counterbalanced lessons. Thus I was able to evaluate transfer effects *between* subjects (rather than within). Sample sizes for this analysis were mostly equal: *pause n*=20, *number n*=20, *title n*=19, *question n*=21.

Design & Procedure

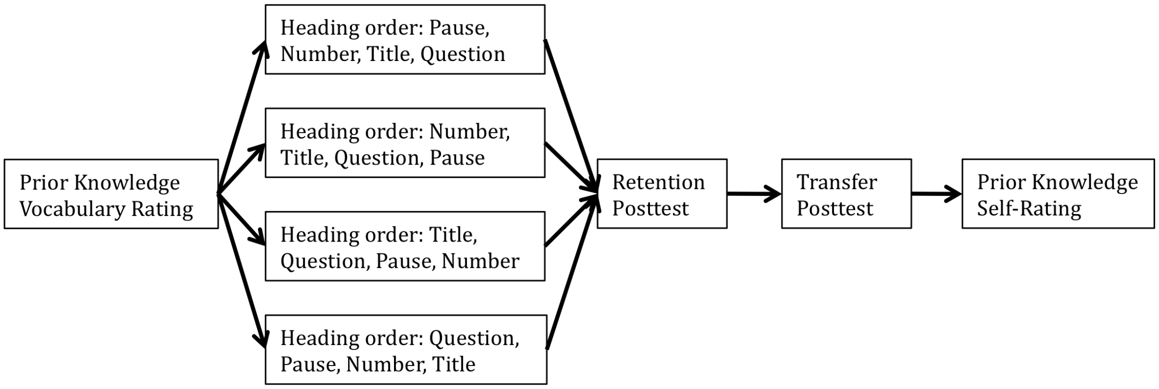


Figure 14. Design and procedure for Experiment 4.

The design and procedure for Experiment 4 are displayed in Figure 14. The study used a one-way within-subjects design, where segment heading was manipulated across segments of the video in 4 conditions: pauses, numbers, titles, or questions. Participants were randomly assigned to one of four video lessons in which the specific order of the headings was counterbalanced across lessons. All participants saw the same instructions:

> "In this video you will learn about the excretory (urinary) system. At the end of the video, you will take a short test to help us evaluate the quality of the lesson. Before you begin, we will ask a few questions to gauge how much you already know about this topic."

After reading the instructions, participants were shown the 10 vocabulary terms and indicated how many of them they could define. They were then shown the following instructions:

> “Please treat this video like something you would watch for fun or to learn something new. This is not a test of your aptitude or knowledge; we are just interested in how much people naturally remember from videos like this.
>
> In order to help you process the lesson, we have broken it down into shorter segments with pauses in between them. When you reach the end of a segment, you may wait for as long as you like before clicking "play" to continue.”

These instructions were followed by the first segment heading (as described above), and then participants watched the video lesson. At the end of the lesson, participants were instructed that they would take a short test on the information presented:

> “Thanks for watching! You will now take a short test on the information in the video and then rate the quality of the video lesson. There are 20 multiple-choice questions on the test. Press play when you are ready to begin.”

Participants then completed both the retention test (16 questions) and the transfer test (4 questions). From the participants’ perspective, the posttest was simply a single test of 20 questions. After completing the transfer test, participants were asked to rate how much they enjoyed watching the video, how much they thought they learned from the lesson, whether they would recommend the lesson to someone else, and whether they would be interested in learning more about this topic. They were also asked to rate how much of the information in the lesson they knew before watching the video; this rating was included as part of the prior knowledge score, as described above. Before finishing the study, participants also reported their age, gender, and highest level of education.

Results

Overall retention scores across all items were non-normally distributed (Shapiro-Wilk: $W(80)=0.961$, $p=0.017$). The mean overall retention score was 0.47 with a standard deviation of 0.20, but the histogram shows two peaks at approximately 0.25 and 0.50. Distributions of the scores for each subset of the retention test (pause, number, title, and question) were also non-normally distributed, although the histograms generally show peaks in the center of the distribution (*pauses* *M*=0.46, *SD*=0.26; *numbers* *M*=0.47, *SD*=0.29; *titles* *M*=0.46, *SD*=0.29; *questions* *M*=0.50, *SD*=0.26). Lastly, transfer was also non-normally distributed (*W*(80)=0.903, *p*<0.001), with a mean of 0.41 and standard deviation of 0.25.

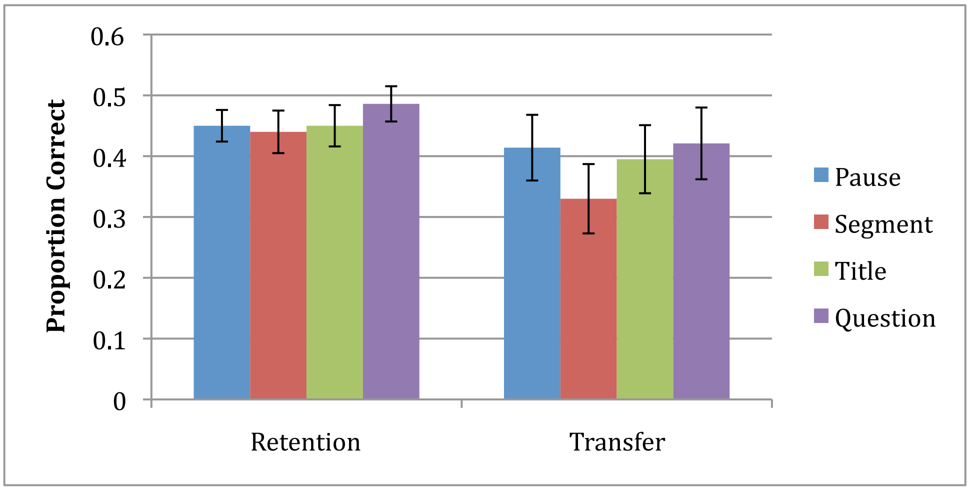


Figure 15. Proportion correct on retention and transfer posttests by condition in Experiment 4. Bars show estimated marginal means with prior knowledge as a covariate (evaluated at PK = 2.96) and error bars indicate standard errors.

Retention was analyzed using a repeated-measures (*heading*: pause, segment, title, question) ANCOVA with prior knowledge as a covariate and video (counterbalanced video conditions) as a between-subjects effect. Estimated marginal means by condition (with prior knowledge as a covariate evaluated at PK =2.96) are shown in Figure 15. The within-subjects effect of heading was not significant (*F*(3,225)=0.932, *p*=0.426). There was also no interaction between prior knowledge and heading (*F*(3,225)=0.747, *p*=0.525) or between video and heading (*F*(9,225)=1.24, *p*=0.274). The between-subjects effect of prior knowledge was significant (*F*(1,75)=10.11, *p*=0.002), but the effect of video was not (*F*(3,75)=0.178, *p*=0.911).

Transfer was analyzed using a one-way between-subjects ANCOVA with prior knowledge as a covariate. The effect of heading was not significant (*F*(3,75)=0.865, *p*=0.463), but the effect of prior knowledge was significant (*F*(1,75)=11.63, *p*=0.001).

Discussion

Experiment 4 did not show a statistically significant effect of heading on either retention or transfer, although the numerical pattern for retention was similar to Experiment 3. *Questions* lead to slightly better retention scores than *titles*, *numbers*, or *pauses*. The pattern of transfer scores in Experiment 4 was somewhat different from the pattern in Experiment 3, but the large standard errors suggest that conditions were not predictably different.

In this experiment the lack of an effect is likely the result of the conditions being somewhat contaminated. Once participants see the variety of headings, they are likely to generate alternative headings than the ones presented. For example, seeing a *numbers* heading like “Segment 3” followed by a *title* heading “Glomerulus and Bowman’s Capsule” followed by a *question* heading “What substances are reabsorbed in the proximal convoluted tubule?” is likely to prompt some participants to come up with alternative headings for subsequent segments. It may also have confused some participants and distracted them from focusing on the lesson as they try to figure out why the headings are so different for different segments. A better test of a within-subjects design should make the differences between headings less obvious so that participants are not distracted or influenced by the differences between conditions.

Experiment 5

In Experiment 4, I suspected that participants might have found the different types of headings distracting rather than helpful. Instead of potentially improving retention or understanding for some segments, the changes in headings may have drawn participants’ attention away from the lesson as they tried to figure out a reason for the changing headings. In this experiment, I sought to reduce this effect by making the conditions less apparent to the participants.

To do this, I reduced the number of conditions to just two: *questions* and *no pauses*. Participants in this study would be unaware of the difference between those two conditions because six of the 12 segments would not be marked with any kind of pause; from the participants' perspective the lesson would appear to have just 6 segments rather than 12. Without a pause or heading to draw their attention, participants would not be expected to direct any extra attention to those segments. On the other hand, the segments marked by questions should draw much more attention because they both pause the video and direct participants to look for specific information. This change to the design should reduce the distracting effect of different headings while creating the maximum difference in attention between conditions. If an effect of heading truly exists, this design should capture the strongest version of that manipulation.

Methods

Participants

Participants were 51 workers on Amazon Mechanical Turk (34 female; M_{age} =39.4 years) who were paid $1 each for participation.

Materials

The video lesson was the same one used in the previous two experiments. The lesson segments were the same as well, but half of the segment headings (and therefore half of the pauses) were eliminated. Each participant watched the whole lesson but only saw a heading for 6 of the 12 segments, so that the lesson appeared to have just 6 segments. Segments without a heading were the *no pause* condition, and segments with a heading always presented the *question* heading for that segment. Two versions of the video lesson were created to counterbalance headings across segments. In one version all the even segments were *question* and all the odd segments were *no pause*; in the other version the headings were reversed.

The test of prior knowledge, retention test, and transfer test were identical to those used in Experiment 4. Due to the design of the segment headings in this study, half of the retention questions (8 out of 16) corresponded to the *no pause* and half to the *question* condition. The four transfer questions all corresponded to one condition or the other, so transfer was assessed between-subjects.

Design & Procedure

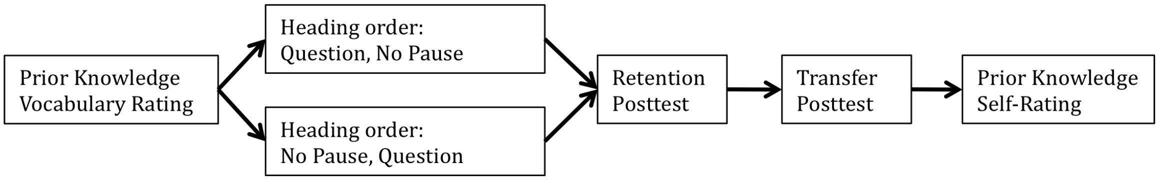


Figure 16. Design and procedure used in Experiment 5.

The design and procedure are displayed in Figure 16. The study used a one-way within-subjects design, where segment heading was manipulated across segments of the video in 2 conditions: no pause or questions. Participants were randomly assigned to one of two counterbalanced versions of the lesson. The procedure was identical to that used in Experiment 4. Participants first completed the prior knowledge measure, and then watched the lesson. Immediately following the lesson, all participants completed the retention and transfer posttests and the rating questions, also reporting their age, gender, and highest level of education.

Results

Overall retention scores (combining both no pause and question subscores) were normally distributed ($W(51)=0.983$, $p=0.650$) with a mean of 0.47 and a standard deviation of 0.20. The *no pause* and *question* subscores were non-normally distributed (no pause: $W(51)=0.949$, $p=0.029$; question: $W(51)=0.941$, $p=0.013$) where the *no pause* distribution ($M=0.43$, $SD=0.23$) was fairly even distributed across the range from 0.1 to 0.8 and the *question* distribution (*M*=0.51, *SD*=0.23) was highly peaked at the center. Transfer was non-normally distributed (*W*(51)=0.889, *p*<0.001) with a mean of 0.47 and a standard deviation of 0.24

Retention scores were first analyzed using a repeated-measures ANCOVA with heading (no pause, question) as a within-subjects factor and prior knowledge as a covariate. In this analysis, the effect of heading was not significant (*F*(1,49)=0.002, *p*=0.963), but the effect of prior knowledge was highly significant (*F*(1,49)=43.92, *p*<0.001) and there was a marginal interaction between heading and prior knowledge (*F*(1,49)=3.45, *p*=0.069).

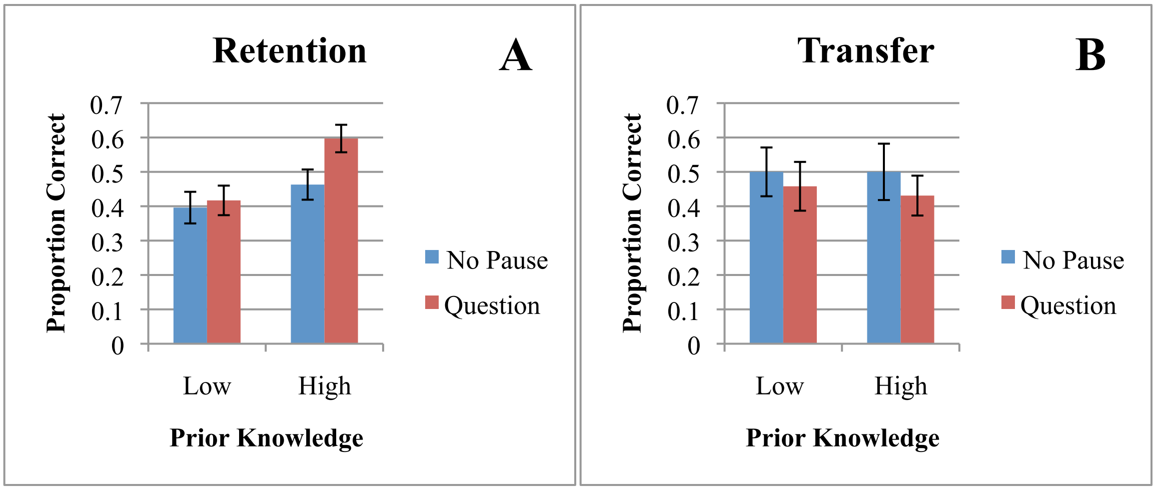


Figure 17. Mean proportion correct on the retention test (A) and transfer test (B) in Experiment 5 for participants with high or low prior knowledge and for segments preceded by a question or no pause. Error bars indicate standard errors.

To follow up on this interaction, I used a median split to separate participants into high- and low-prior knowledge groups. Participants with prior knowledge scores of 1-3 were low knowledge, and participants with scores of 4-15 were high knowledge. I then analyzed retention scores with a mixed ANOVA with heading as a within-subjects factor and prior knowledge split as a between subjects factor. Mean proportion correct on the retention test by prior knowledge (high, low) and heading (no pause, question) is shown in Figure 17A. In this analysis there was a main effect of both heading (*F*(1,49)=7.66, *p*=0.008) and prior knowledge split (*F*(1,49)=5.171, *p*=0.027), as well as a significant interaction (*F*(1,49)=4.10, *p*=0.048). Participants with low prior knowledge did not differ on retention for question and no pause segments (*t*(23)=-0.451, *p*=0.656), but participants with high prior knowledge showed a significant benefit on segments preceded by questions compared to those with no pause (*t*(26)=-4.301, *p*<0.001).

Transfer scores were analyzed using a two-way ANOVA where heading and prior knowledge split were defined between subjects. Mean proportion correct on the transfer test by prior knowledge (high, low) and heading (no pause, question) is shown in Figure 17B. In this analysis, the main effect of heading was not significant (*F*(1,47)=16.00, *p*=0.156), nor was the main effect of prior knowledge split (*F*(1,47)=1.00, *p*=0.500) or the interaction (*F*(1,47)=0.038, *p*=0.845).

Discussion

As predicted, there was a significant effect of heading on retention where retention was higher for *question* segments than for *no pause* segments. However, the interaction with prior knowledge was somewhat unexpected. The results showed that only participants with some familiarity with relevant vocabulary (high prior knowledge) benefitted from the questions, while participants with low familiarity (low prior knowledge) did not benefit at all from the question headings. Perhaps this is because participants with low prior knowledge were overwhelmed by headings that contained terms with which they were unfamiliar, and so were unable to use those headings to guide their attention during the lesson.

Alternatively, scoring high on the measure of prior knowledge could be an indicator of more than just familiarity with the topic and the vocabulary; it is possible, for example, that participants who score high on prior knowledge tend to be better students than those who score low, and thus were more attentive or motivated. This explanation seems less likely because there was no benefit of high prior knowledge for the *no pause* condition, but I cannot rule out the possibility that prior knowledge may have been confounded with other important differences between participants.

The finding that headings might benefit participants with high prior knowledge more than participants with low prior knowledge could be an interesting contribution to Mayer's work on the signaling principle. However, retrospective analyses of Experiments 3 and 4 did not find a similar pattern. Although prior knowledge is strongly related to performance on both retention and transfer, the effect of heading did not interact with prior knowledge (Experiment 3: $F(3,74)=0.135$, $p=0.939$; Experiment 4: $F(3,234)=2.10$, $p=0.101$). Prior knowledge seems to play a role in the outcomes of these experiments, but it's not exactly clear how large that effect is and how it interacts with the heading manipulation.

One potential issue with the studies run so far is that most Mechanical Turk participants are not students and may have been out of the classroom for a long time. It may be that students who are currently in school interact with instructional videos in a different way than adults who are not in school. For example, undergraduates may approach the video with different goals for learning – such as focusing on details or gist understanding – at different rates than non-students. Also, there is no way to know how seriously the participants were engaging with the lesson. Popular articles about Amazon's Mechanical Turk suggest that workers are often multi-tasking while doing Turk jobs. This could seriously impact our ability to tell whether this manipulation affects learning for students who are actually engaged in a course. A more controlled setting with an experimenter in the room and undergraduate students as subjects might reveal a different pattern of results than I have found over Mechanical Turk.

Experiment 6

In Experiments 3, 4, and 5, several manipulations of segment headings – both between- and within-subjects – yielded mixed results. One factor that may have contributed to this was the use of Amazon Mechanical Turk to recruit participants. Mturk workers are generally not students, and there is very little control over how they engage with the video lesson. It is possible that the participants in the last three experiments may have been distracted while completing the task or might have been less motivated to take the study seriously. This experiment addressed those issues by recruiting UCLA undergraduates and by asking them to come into the lab to participate in the study. Although the entire study was completed online, participants were seated in a room with the experimenter present, which should have motivated them to take the lesson and the posttest more seriously.

Additionally, this experiment provides a stronger test of the segmenting principle. In Experiments 3 and 4 the primary manipulation was the type of heading (i.e. signal) that was provided at the beginning of each segment, but neither experiment tested the benefits of the pauses (i.e. segments) themselves. Experiment 5 included a condition without pauses, but did not have a strong control for the segmenting principle given that the *question* condition was both segmented and signaled. In this experiment I have a stronger test of the segmenting principle by including both a condition with no pauses as well as a condition with pauses but no headings.

Given that Experiments 3 and 4 showed minimal differences between the *pauses* and *numbers* conditions, I decided to eliminate the *numbers* condition. Thus this experiment had four conditions total: *no pauses*, *pauses*, *titles*, and *questions*. These conditions were manipulated between subjects because I was concerned that the within-subjects design would lead to too much contamination between conditions. I predicted that the *no pauses* condition would perform worst overall since it did was neither segmented nor signaled. I expected the *pauses* condition to perform slightly better than *no pauses* because the pauses would allow students to self-pace. Lastly, I predicted that the *titles* and *questions* conditions would perform best because they were both segmented and signaled, though the *questions* condition should score highest overall because the headings were more specific and directed.

Methods

Participants

Participants were 95 undergraduates (60 female; M_{age} =20.1 years) who participated for course credit.

Materials

Video lesson. The video lesson and segments were the same as those used in the previous three experiments. Four versions of the video lesson were created in which all 12 headings were of the same type. *Pause*, *title*, and *question* headings were identical to those used in Experiments 3 and 4. The *no pause* condition simply eliminated all headings and thus let the video play uninterrupted.

Posttest. The test of prior knowledge was identical to the one used in Experiments 4 and 5, but changes were made to both the retention and transfer tests. The complete retention test can be found in Appendix F. Both the retention and transfer test included one open response questions and 12 multiple-choice questions, one corresponding to each lesson segment. The retention open response question asked participants to describe how the kidney filters blood, using as much detail as possible. A word bank was provided to help participants remember and use key vocabulary from the lesson. Answers were scored on a scale from 1-5 according to the detail and accuracy of the answers. The average score on across all participants was 2.3 (SD=1.51).

Retention multiple-choice questions required participants to recognize information that was presented directly in the lesson.

The transfer open response question asked whether water, a sports drink, or a soda would rehydrate a dehydrated person more quickly, and why. Answers were scored on a scale from 1-5 according to the validity of the answers and the amount of supporting detail provided. The average score across all participants was 3.0 (SD=0.84). Transfer multiple-choice questions asked participants to apply information from the lesson to make predictions, troubleshoot, or choose the best explanation for various scenarios related to kidney function.

Design & Procedure

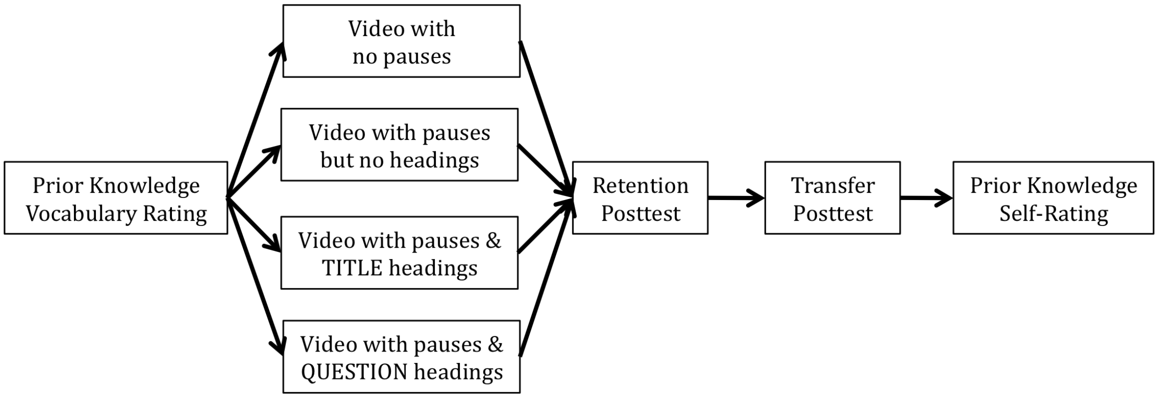


Figure 18. Design and procedure used in Experiment 6.

The design for Experiment 6 is shown in Figure 18. The procedure was identical to that used in Experiment 3. Participants were randomly assigned to one of 4 between-subjects conditions: *no pauses* (*n*=23), *pauses* (*n*=24), *titles* (*n*=24), or *questions* (*n*=24). They first completed the prior knowledge measure, and then watched the video lesson. Immediately following the lesson, all participants completed the posttest and rating questions, also reporting their age, gender, and highest level of education. One additional question was added during the posttest, asking participants what they did during the pauses in the lesson (this question was omitted for the *no pause* condition).

Results

Both retention and transfer scores were normally distributed. Overall, the mean retention score was 0.59 with a standard deviation of 0.22 ($W(95)=0.980$, $p=0.157$). The mean transfer score was 0.54 with a standard deviation of 0.15 ($W(95)=0.978$, $p=0.114$).

The effect of heading on retention and transfer was analyzed using a one-way MANCOVA with heading (4: no pause, pause, title, questions) as a between-subjects factor and prior knowledge as a covariate. There was no significant effect of heading on either retention ($F(3,90)=1.55$, $p=0.207$) or transfer ($F(3,90)=0.782$, $p=0.507$), but there was a significant effect of prior knowledge on both retention ($F(1,90)=47.44$, $p<0.001$) and transfer ($F(1,90)=11.25$, $p=0.001$).

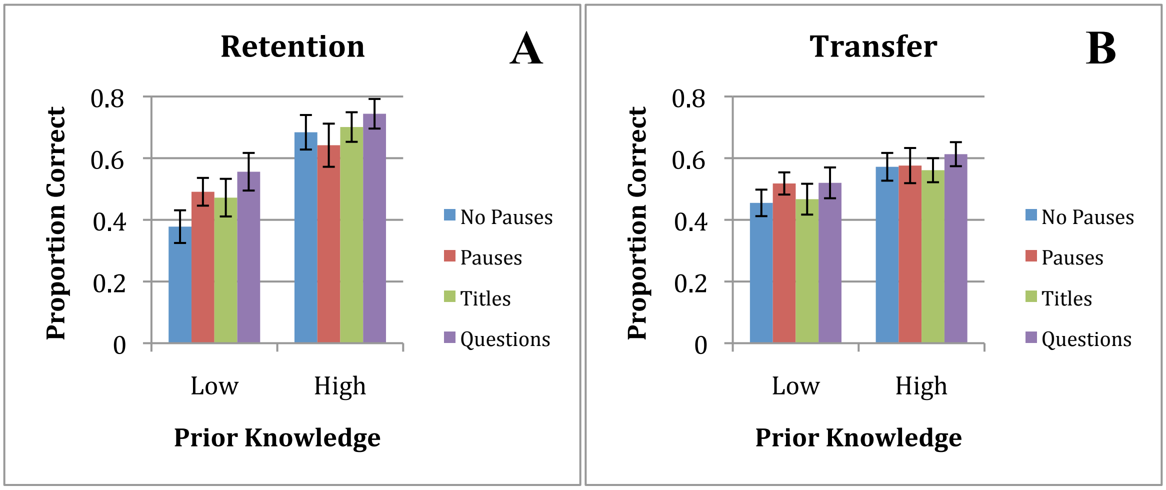


Figure 19. Mean proportion correct on the retention test (A) and transfer test (B) in Experiment 6 by type of heading and high or low prior knowledge. Error bars indicate standard errors.

To test for an interaction between prior knowledge and heading, I used a median split on prior knowledge to split participants into high- and low-prior knowledge groups, and used a two-way MANOVA with heading and prior knowledge split as between-subjects factors. The estimated marginal means for retention and transfer by heading and prior knowledge split are shown in Figure 19. This analysis showed a significant main effect of prior knowledge split on both retention (*F*(1,87)=30.65, *p*<0.001) and transfer (*F*(1,87)=7.93, *p*=0.006), but no main effect of heading (Retention: *F*(3,87)=1.64, *p*=0.186; Transfer: *F*(3,87)=0.673, *p*=0.571) and no interaction between prior knowledge split and heading (Retention: *F*(3,87)=0.71, *p*=0.549; Transfer: *F*(3,87)=0.14, *p*=0.936).

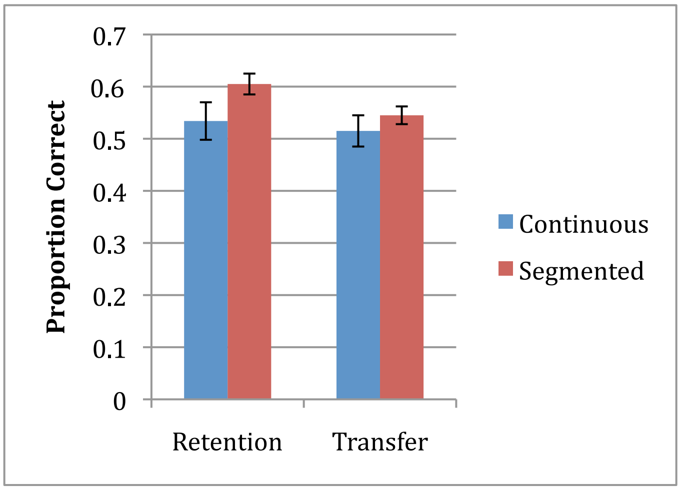


Figure 20. Mean proportion correct on retention and transfer in Experiment 6 for continuous and segmented video lessons. Error bars indicate standard errors.

Although the effects of heading on retention and transfer were not significant, I separately analyzed the effects of segmenting and signaling. To analyze segmenting, I collapsed the three conditions with pauses (*pauses*, *titles*, and *questions*) and compared them against the *no pause* condition using a one-way MANOVA with retention and transfer as outcomes, segmenting (2: segmented, continuous) as a between-subjects factor, and prior knowledge as a covariate. The mean proportion correct in these groups is shown in Figure 20. The effect of segmenting was marginally significant for retention (*F*(1,92)=2.91, *p*=0.091) but was not significant for transfer (*F*(1,92)=0.758, *p*=0.386).

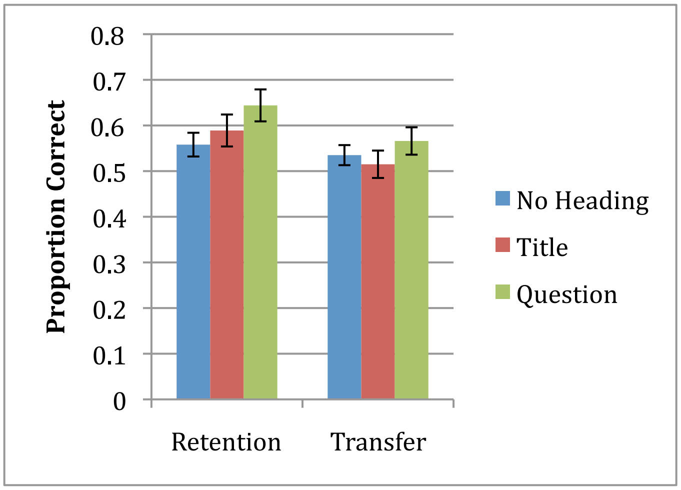


Figure 21. Mean proportion correct on retention and transfer in Experiment 6 for video lessons with no headings, title headings, and question headings. Error bars indicate standard errors.

To evaluate the effect of signaling, I collapsed the two non-signaled conditions (*no pauses* and *pauses*) and compared them against the *title* and *question* conditions using a one-way MANOVA with retention and transfer as outcomes, signaling (3: no heading, titles, questions) as a between-subjects factor, and prior knowledge as a covariate. The mean proportion correct in these groups is shown in Figure 21. The effect of signaling was not significant for either retention ($F(2,91)=1.86$, $p=0.161$) or transfer ($F(2,91)=0.743$, $p=0.479$).

	Activity During Pauses				
Condition	Play	Review	Wander	Mixed	TOTAL
Pauses	19	2	1	1	23
Titles	19	4	0	1	24
Questions	12	8	2	2	24
TOTAL	50	14	3	4	71

Table 3. Frequency of reported activity during pauses by condition in Experiment 6.

When asked what they did during the pauses betweens segments, participants' responses could be categorized in one of 4 mutually exclusive ways. One participant was excluded from this analysis for not providing a response to this question. A majority of participants (70.4%) responded that they clicked play right away ("Play"). Fewer participants (19.7%) reported trying to review the information from the previous segment before moving on ("Review"). Just 3 participants (4.2%) said they let their minds wander during the pauses ("Wander"). The remaining 4 participants (5.6%) reported a mix of strategies ("Mixed"). Generally they started out trying to review information but after a few pauses felt overwhelmed and gave up. The rate at which participants reported these types of activities by condition can be seen in Table 3 above, but did not differ significantly ($X^2(6)=8.40$, $p=0.210$).

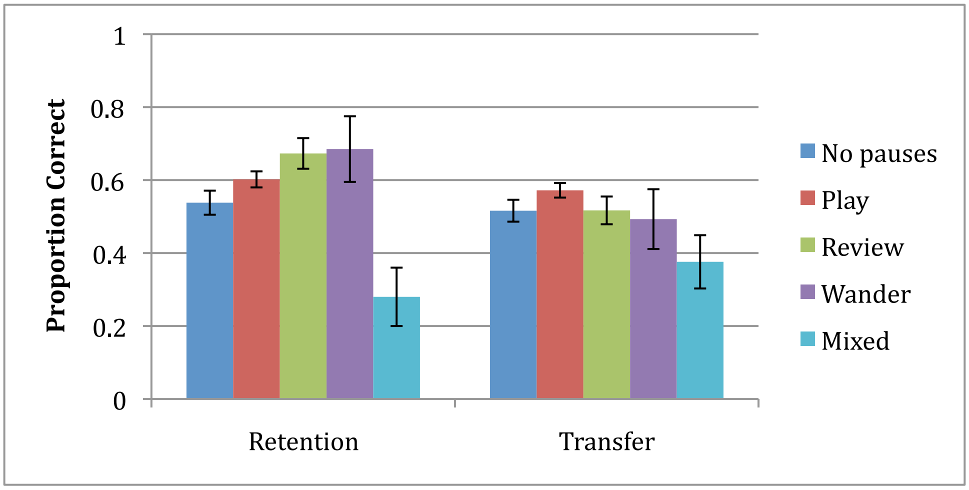


Figure 22. Mean proportion correct on retention and transfer in Experiment 6 as a function of the activity participants engaged in during video pauses. Error bars indicate standard errors.

Although the conditions did not differ in the rate at which participants reported reviewing the material, I suspected that what participants did during the pauses would affect their performance on the retention and transfer tests. I tested this prediction using a one-way MANCOVA with retention and transfer as outcomes, prior knowledge score as a covariate, and pause activity (5: no pauses, play, review, wander, mixed) as a between-subjects factor. The mean proportion correct on retention and transfer for these groups is shown in Figure 22. The effect of pause activity was statistically significant for retention ($F(4,88)=5.66$, $p<0.001$) and was marginally significant for transfer ($F(4,88)=2.24$, $p=0.071$). Post hoc analyses of retention showed that the mixed activity scored significantly lower than all other groups, but the other groups (no pauses, play, review, and wander) were not different from each other, although the numerical pattern shows an advantage for reviewing and mind-wandering. Post hoc analyses of transfer showed that no two groups were statistically different from each other, although the numerical pattern shows an advantage for clicking play right away.

Discussion

This experiment again failed to find a significant effect of heading on either retention or transfer. Numerically we continue to see a trend of higher scores for question headings, lower scores for no headings, and intermediate scores for title headings. When analyzed as separate factors, segmenting had a marginal effect on retention but no effect on transfer, while signaling did not have a significant effect on either retention or transfer. These results suggest that segmenting may provide at least some benefit, but the effect is not very large. A likely explanation for this is the somewhat low power of the study; with only 24 participants per condition, it is difficult to detect what might be a fairly small effect.

The analysis of participants' reported activity during the pauses might provide more insight into the effect of the pauses and headings. Although the rate at which participants reported clicking play, reviewing, mind wandering, or a mix of strategies did not differ significantly between conditions, the lack of significance was likely due to a small sample size. The frequency of reviewing in the *questions* condition was double that reported in the *titles* condition and four times the frequency in the *pauses* condition. The increase in the rate of reviewing was accompanied by a decrease in the rate of simply clicking play. These differences in activity likely contributed to the slight advantage of the *questions* condition on the retention test because reviewing during the pauses lead to higher retention than clicking play right away. The mixed activity group performed worst overall, but most of those participants reported giving up on reviewing because they felt overwhelmed. Thus their low performance is likely a result of reduced motivation rather than an ineffective combination of strategies.

These findings suggest that the effect of signaling may be confounded to some extent with what participants are doing during the pauses. In one sense this is not a confound, but rather the expected effect of different kinds of headings. However, the tendency of most participants to click play without reviewing material implies that the headings are not effective because participants are not engaging with them.

Another possible reason that the effect of signaling was non-significant could be that the posttest was given immediately after watching the lesson. Some robust memory phenomena such as the testing effect (Roediger & Karpicke, 2006) only show effects at a delay. If the effect of headings is similar, I may not be able to detect a difference between conditions at immediate posttest. Although there is not strong previous evidence that a delayed posttest would enhance the effect of signaling, it is nonetheless possible that signals may benefit learning by helping participants better integrate the lesson with their prior knowledge, thus facilitating later retrieval.

Experiment 7

Experiment 6 failed to find an effect of heading on retention or transfer when the participants were UCLA undergraduates. The analyses showed a small benefit of segmenting for retention but not for transfer, and no benefit of signaling for either retention or transfer. A possible explanation for the lack of a significant effect is that the benefits of signaling may be more pronounced at a delay. Experiment 7 addresses this hypothesis by introducing a 24-hour delay between watching the video and completing the posttest. If the delay enhances the effect of the headings, I would make the same predictions about the relative order of performance for the heading conditions, but I would expect overall performance to be lower than in Experiment 6. If the delayed posttest does not lead to a significant effect of heading, then it is much more likely that such an effect does not exist or is being affected by other extraneous variables.

Methods

Participants

Participants were 102 undergraduates (73 female; M_{age} =20.7 years) who participated for course credit.

Materials

The materials in this study were identical to those used in Experiment 6. The only major difference was that the posttest was presented using Collector, an open-source PHP-based program designed to run psychology experiments and conducted via an Internet browser (https://github.com/gikeymarcia/Collector). Using this program allowed me to randomize the order of the multiple-choice questions on the posttest. In the four previous studies, all questions were always presented in the same order to all subjects.

In this study, the open response retention question was always presented first, followed by the 12 multiple-choice retention questions in a random order, followed by the 12 multiple-choice transfer questions in a random order, and lastly followed by the open response transfer question. I used this order so that the retention open response would not be influenced by the content of the multiple-choice questions; I wanted the open response answer to reflect what students remembered from the lesson without reminding. The open response transfer question was intentionally last so that students would have the benefit of having seen all the multiple-choice questions first. Transfer questions tend to be more difficult to answer than retention questions, and I hoped that taking the multiple-choice test would reactivate information that would support their ability to answer that question thoroughly and correctly.

Design & Procedure

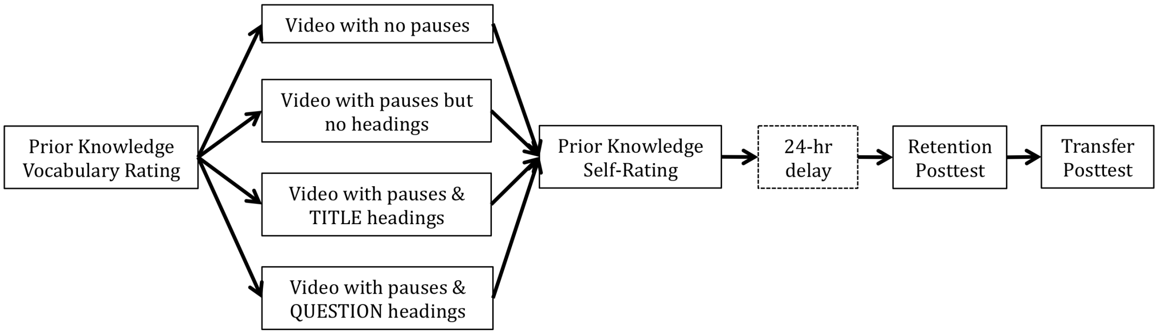


Figure 23. Design and procedure used in Experiment 7.

The design and procedure in Experiment 7 are displayed in Figure 23. Participants were randomly assigned to one of 4 between-subjects conditions: no pauses (*n*=25), pauses (*n*=25), titles (*n*=26), or questions (*n*=26). They first completed the prior knowledge measure, and then watched the lesson and completed the rating questions (including the question asking what they did during the pauses), also reporting their age, gender, and major. The next day, participants returned to the lab and completed both the retention and transfer posttests. The procedure was identical to that used in Experiment 6 except that participants completed the rating questions immediately after watching the video, but completed posttest 24 hours after watching the lesson.

Results

Overall, retention scores and transfer scores were distributed normally. Retention scores had a mean of 0.47 and a standard deviation of 0.17 ($W(102)=0.980$, $p=0.128$). Transfer scores had a mean of 0.48 and a standard deviation of 0.16 ($W(102)=0.989$, $p=0.555$). These scores were somewhat lower than the overall scores in Experiment 6 (Retention $M=0.59$; Transfer $M=0.54$). The effect of heading (4: no pauses, pauses, titles, questions) on retention and transfer scores was first analyzed using a one-way MANCOVA with prior knowledge score as a covariate. The effect of heading was not significant for either retention ($F(3,97)=1.615$, $p=0.191$)

or transfer (*F*(3,97)=0.385, *p*=0.764). Prior knowledge score, however, was significantly related to both retention (*F*(1,97)=57.72, *p*<0.001) and transfer (*F*(1,97)=21.25, *p*<0.001).

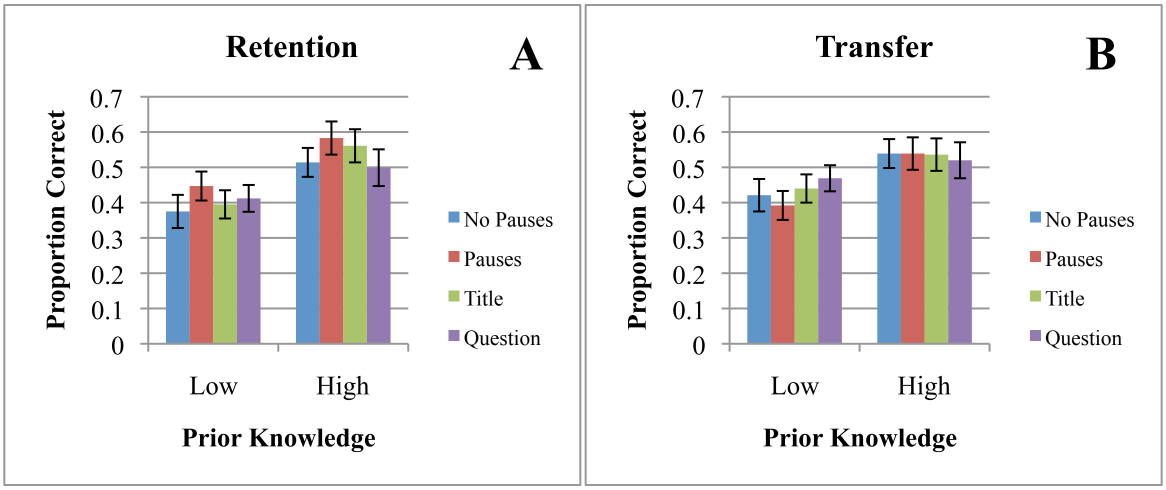


Figure 24. Mean proportion correct on the retention test (A) and transfer test (B) in Experiment 7 by type of heading and high or low prior knowledge. Error bars indicate standard errors.

A follow-up analysis used a median-split to separate participants into high- and low-prior knowledge groups and then used a two-way MANOVA to assess the interaction between heading (4: no pause, pause, title, question) and prior knowledge split (2: low, high). Low prior knowledge was a score of 4 or lower on the prior knowledge composite score (*n*=57), while high prior knowledge was a score of 5 or higher (*n*=45). Mean proportion correct on retention and transfer by prior knowledge split and heading is shown in Figure 24. The effect of prior knowledge split was still highly significant for both retention (*F*(1,94)=17.79, *p*<0.001) and transfer (*F*(1,94)=11.03, *p*=0.001), but neither interaction was significant (Retention: *F*(3,94)=0.273, *p*=0.845; Transfer: *F*(3,94)=0.420, *p*=0.739), nor was either main effect of heading (Retention: *F*(3,94)=0.992, *p*=0.400; Transfer: *F*(3,94)=0.162, *p*=0.922).

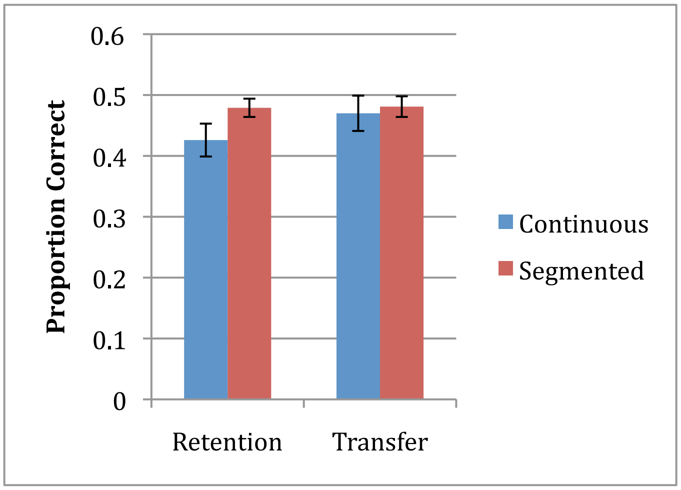


Figure 25. Mean proportion correct on retention and transfer posttests in Experiment 7 for continuous and segmented video lessons. Error bars indicate standard errors.

To separately analyze the effect of segmenting, the three conditions with pauses (*pauses, titles, questions*) were collapsed and compared against the *no pause* condition. Then the effect of signaling was analyzed using a one-way (2: continuous, segmented) MANCOVA with retention and transfer as outcomes and prior knowledge as a covariate. The mean proportion correct on retention and transfer for continuous and segmented video lessons is shown in Figure 25. There was a marginally significant effect of segmenting on retention (*F*(1,99)=3.01, *p*=0.086) but no effect of segmenting on transfer (*F*(1,99)=0.106, *p*=0.745). The effect of prior knowledge was highly significant for both retention (*F*(1,99)=59.11, *p*<0.001) and transfer (*F*(1,99)=21.00, *p*<0.001). A follow-up analysis with prior knowledge split as a between-subjects factor showed no interaction between prior knowledge split and retention (*F*(1,98)=0.004, *p*=0.981) or transfer (*F*(1,99)=0.098, *p*=0.756).

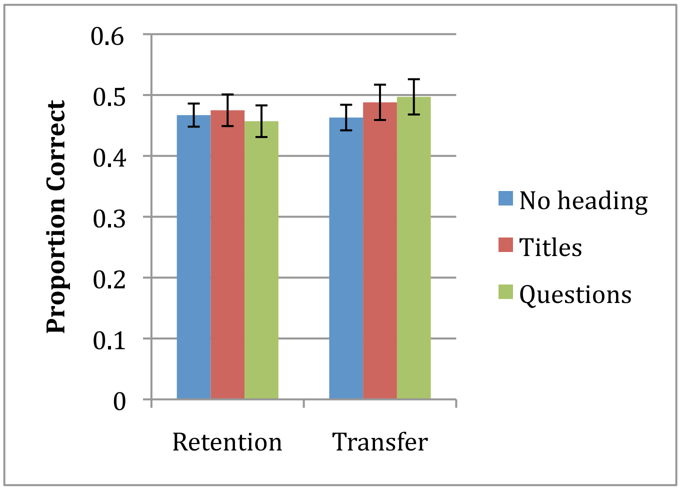


Figure 26. Mean proportion correct on retention and transfer in Experiment 7 for no headings, title headings, and question headings. Error bars indicate standard errors.

To assess the effect of signaling, the two conditions without headings (*pauses* and *no pauses*) were collapsed and compared with the *title* and *question* conditions using a one-way (3: no heading, title, question) MANCOVA with retention and transfer as outcomes and prior knowledge as a covariate. The mean proportion correct on retention and transfer for the no heading, title heading, and question heading conditions is shown in Figure 26. The effect of signaling was not significant on either retention ($F(2,98)=0.119$, $p=0.888$) or transfer ($F(2,98)=0.536$, $p=0.587$), but the effect of prior knowledge was highly significant on both retention ($F(1,98)=53.25$, $p<0.001$) and transfer ($F(1,98)=21.90$, $p<0.001$). A follow-up analysis with prior knowledge split as a between-subjects factor showed no interaction between prior knowledge split and retention ($F(2,96)=0.397$, $p=0.674$) or transfer ($F(2,96)=0.619$, $p=0.541$).

	Activity During Pauses				
Condition	Play	Review	Wander	Mixed	Total
Pauses	18	4	0	3	25
Title	22	3	1	0	26
Question	13	8	1	4	26
TOTAL	53	15	2	7	77

Table 4. Frequency of reported activity during pauses by condition in Experiment 7.

When asked what they did during the pauses betweens segments, participants' responses were categorized in the same was in Experiment 7. A majority of participants (68.8%) responded that they clicked play right away. Roughly one fifth of participants (19.5%) reported trying to review the information from the previous segment. Two participants (2.6%) said they let their minds wander during the pauses. The remaining 7 participants (9.1%) reported a mix of strategies. Similar to the reports in Experiment 6, participants in the mixed strategy category said they tried to review but gave up when they felt overwhelmed. The rate at which participants reported these types of activities by condition can be seen in Table 4 above, but did not differ significantly by condition ($X^2(6)=9.71$, $p=0.138$).

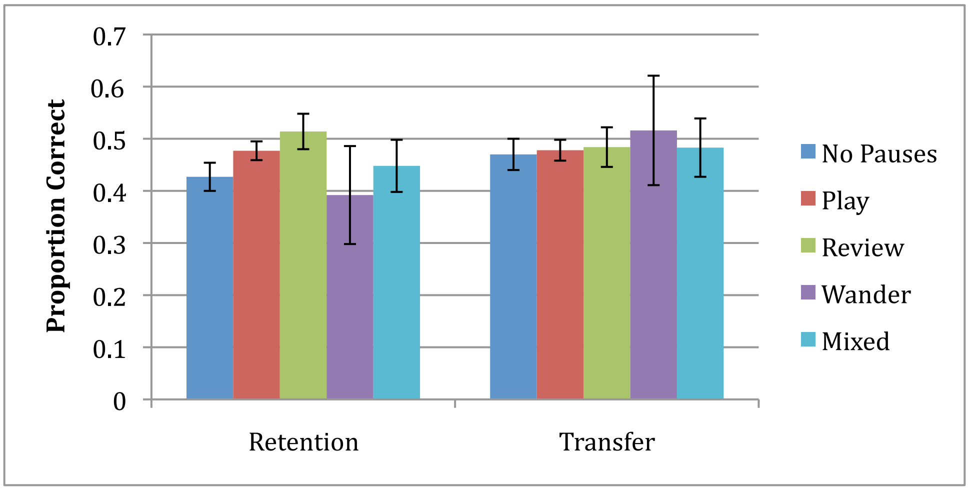


Figure 27. Mean proportion correct on retention and transfer in Experiment 7 as a function of the activity participants engaged in during video pauses. Error bars indicate standard errors.

Retention and transfer performance as a function of the reported activity was analyzed using a one-way MANCOVA with prior knowledge as a covariate. The mean proportion correct on both posttests as a function of pause activity (5: no pauses, play, review, wander, mixed) is shown in Figure 27. The effect of pause activity was not significant for either retention ($F(4,96)=1.32$, $p=0.269$) or transfer ($F(4,96)=0.061$, $p=0.993$).

Discussion

Rather than amplifying the effect of heading, the delayed posttest seemed to diminish it. Overall, scores in this experiment were much lower than those in Experiment 6, suggesting that the delayed posttest depressed performance overall. However, the previous pattern of questions outscoring titles and pauses was eliminated or (in some cases) reversed. In this study, there appears to be a numerical advantage for the *pauses* group over *titles*, and of *titles* over *questions*. Given the lack of statistical significance, however, I hesitate to interpret what this could mean. If the pattern is replicated in further experiments using delayed posttests, it could indicate that providing segment headings benefits immediate test performance but hurts delayed performance. However, at this point, it is also possible that the results are due to chance variation.

Unlike in Experiment 6, the activity participants engaged in during the pauses did not significantly affect performance on retention or transfer. The pattern of results still suggests and an advantage of reviewing material for the retention posttest, but there was no observable difference in transfer performance by pause activity. It may be that the delayed posttest diminished the effect of those activities because any resultant learning from review was lost over the retention interval.

General Discussion

In this series of experiments, section headings seemed to do little to improve participants' learning from the lesson. In four of the five experiments there appeared to be a *slight* numerical advantage for the *question* headings on the retention test (and occasionally on the transfer test, as well), but the effect was only statistically significant in one experiment. Some experiments suggested that headings might benefit only participants with high prior knowledge, but this pattern was not consistent and other experiments suggested the opposite: that headings have a greater benefit for participants with *low* prior knowledge. Explicit analysis of the separate effects of segmenting and signaling in Experiments 6 and 7 suggested that segmenting the lesson lead to a small improvement in retention (marginally significant in both studies) but had no effect on transfer. The effect of signaling was not significant for retention or transfer in either experiment, although the numerical trend still suggests a possible advantage for question headings. Delayed posttests diminished, rather than enhanced, this effect.

Such mixed results do little to clarify the usefulness of segmenting or signaling for improving students' learning from video. They do, however, suggest that Mayer's (2008) work is not straightforwardly applicable to all kinds of multimedia lessons. There are important differences between the materials used in Mayer's work and those used here. For example, in the studies of segmenting (Mayer & Chandler, 2001; Mayer, Dow, & Mayer, 2003) the multimedia lesson was under 3 minutes long and each segment contained just one or two sentences of narration with 8-10 seconds of animation. His lessons also did not feature an on-screen narrator. Rather, they use voiceover narration with animated drawings. Lastly, his lessons tend to teach relatively simple process concepts that involve just a few steps.

The lesson used in these studies, however, was just over 10 minutes long, and the segments ranged in length from 23 to 103 seconds. It covered just one "system" in the human body, but addressed several different processes that involved many steps and lots of new vocabulary to describe accurately. This lesson also combined the use of animations, on-screen text, voiceover narration, and an on-screen narrator. In short, the lesson used in this study was much longer and more complex (both conceptually and audio-visually) than the lessons Mayer used to demonstrate the benefits of signaling and segmenting.

The failure to replicate the effects of segmenting and signaling with the present materials does not imply that those principles of multimedia instruction do not exist, but rather that the application of those principles is more complicated than existing work suggests. Much more work is needed to understand the conditions under which segmenting and signaling (and perhaps other principles of multimedia instruction) benefit student learning, as well as the conditions under which learning may be unaffected or perhaps decreased. It will be important to consider such factors as the topic of instruction, the length and complexity of the video, and the characteristics of the learners that might affect the effectiveness of the lesson.

Limitations

One major limitation of this study was that I did not include a condition in which participants took the posttest without watching the video. Thus, I cannot say with certainty that participants actually learned anything from the video. I suspect that they did learn, and self-reports (in response to the question "How much do you think you learned from this video?") indicated that participants believed they had learned something (the average self-rating across all experiments was 3.2, *SD*=1.00). But without a measure of pre-video performance it is hard to be sure of how much students truly learned, if they learned at all.

Another limitation of this study was the low engagement with the material. Evidence from Experiments 6 and 7 suggests that participants did not take advantage of the pauses in the video to think about or review information; rather they simply clicked play. Some participants even reported that they found the pauses annoying or distracting. It should not be surprising then that the headings had a little effect on learning; if students do not pause to read or think about them, they may as well not be presented. The analysis of reported activity during the pauses lends further weight to this interpretation because it suggests that students who *did* take time to pause and review the lesson retained more of it than students who moved on right away.

Future Directions

One other possible flaw of the study was that headings were not visible during the entire video. When headings have been used to improve the global coherence of text passages, they are generally visible while participants read those passages. In this study, however, the heading was shown before and after the lesson segments, but was not visible during the lesson. This adds an extra processing burden because remembering the headings would occupy some working memory resources. Future studies of segment headings could compare the type of design used in these experiments against video lessons with headings visible throughout to determine whether this could have played a role in the lack of an effect.

Another avenue to explore involves encouraging students to engage more deeply with the material presented during the pauses. Posing questions (rather than providing headings) might accomplish this, especially if students are required to respond in some way. Future studies might assess the effects of different types of questions (multiple-choice, open response, etc.) on both the coherence of the lesson and later retention and transfer. Studies from the text comprehension literature suggest that interpolated questions can be effective for increasing comprehension of written texts (Rickards, 1976; Rickards & DiVesta, 1974; Rickards & Hatcher, 1977; Rothkopf, 1966; Rothkopf & Biscibos, 1967). This may be a more fruitful path for future research on coherence and comprehension of video lessons.

Chapter 5: Effects of Advance Organizers on Recall and Transfer

The goal of this study is to evaluate what kinds of materials or activities presented *before* a video lesson will affect participants' learning. Evidence from both studies of text comprehension and multimedia learning suggests that prior activities can change what participants understand and remember from a lesson, and a major priority of this study is to replicate those findings.

One particular manipulation of interest concerns whether the advance organizer is consistent or inconsistent with the organization of the lesson itself. This was the key manipulation in a study by Mannes and Kintsch (1987), which revealed that consistent outlines benefitted participants' retention of details from the text, but inconsistent outlines aided their ability to make inferences and solve problems related to the text. In that study, participants read an article about the history of bacteria used in industrial processes. Prior to reading the article, all participants studied an outline that contained information relevant to the text as well as general information about bacteria. The experimental difference between conditions was whether the outlines were organized consistently with the article (i.e. with similar section headings in the same order) or inconsistently with the article. Critically, all participants studied all the same information; the only difference was its organization. Additionally, none of the information in the outlines could be used directly to answer the posttest questions.

The interpretation of Mannes and Kintsch's (1987) results suggested that consistent outlines benefitted retention because they provided readers with advance organization of the text itself. Participants who studied the consistent outline saw a conceptual structure that could be fleshed out by the article without having to restructure any of the preceding information. In the inconsistent condition, however, the outline provided relevant information but participants had to reorganize it to connect with the information in the text. For this reason, Mannes and Kintsch argued that the inconsistent outlines provided readers with greater prior knowledge but did not enhance the structure of the text base. Therefore, participants' memory of specific details was weaker, but their ability to make inferences about the text was improved because they were building more connections between ideas, rather than filling slots in a structure.

Experiment 8

This experiment attempts to replicate Mannes and Kintsch's (1987) study using video lessons instead of text passages, but also to test a fully crossed design. In the original study, two sources of information were used to create the outlines – the article to be learned and an encyclopedia entry relevant to the article – but participants did not ever study the encyclopedia article. Thus it was not possible to know whether the critical difference between conditions was truly the consistency of organization, or some other quality of the outlines. In this study, I used two video lessons on the same topic so that outlines consistent with each video lesson could be crossed with the other video, thus creating two consistent and two inconsistent conditions. I also included two conditions in which participant did not study an outline at all, so that I could isolate the effects of studying any outline at all.

A single posttest was designed so that it could be answered after having watched either video lesson, so all participants took the same final test. The posttest included both retention and transfer items designed to measure free recall, cued recall, and application of key ideas to novel scenarios. If the outlines function as advance organizers, then we should expect that participants who study an outline before watching the video would perform better overall than participants who only watch the video. If the original finding from Mannes and Kintsch (1987) is replicated, then we should also expect participants who study consistent outlines to perform on the retention test than participants who study inconsistent outlines, but the reverse should be true for performance on the transfer test.

Methods

Participants

Participants were 86 undergraduates (69 female; M_{age} =20.9 years) who participated for course credit.

Materials

Video lessons. The video lessons used in this study were two short lessons about human memory. The first video was produced by Crash Course, a popular YouTube channel that creates 10-minute videos covering topics typical of introductory courses in various domains. This video was 9 minutes and 19 seconds long and covered several topics relevant to an introductory discussion of human memory, including Clive Wearing, types of memory tests (recall, recognition, relearning), the Atkinson-Shiffrin model (sensory memory, STM, LTM), working memory, types of long-term memory (semantic, episodic), and tips for improving your memory (mnemonics, levels of processing). The Crash Course video uses a combination of animations, text slides, and an on-screen narrator to convey the lesson material. The video is fast-paced and entertaining, weaving in verbal and visual jokes along with the lesson content. The original YouTube video can be found at the following link: https://www.youtube.com/watch?v=bSycdIx-C48

The second video was produced by Khan Academy, a YouTube channel well known for creating review materials on a variety of topics. In contrast to Crash Course, Khan Academy videos rely exclusively on tablet writing and drawing with voiceover narration. This video is somewhat slower paced, and all written or drawn information stays on screen for the entire video. The Khan Academy lesson was 7 minutes and 33 seconds long, and covered similar topics to the Crash Course lesson: the information processing model of memory, sensory memory (iconic, echoic), working memory (visuospatial sketchpad, phonological loop, central executive, episodic buffer), and long-term memory (including distinctions between explicit/implicit, semantic/episodic, and procedural/priming memory). The original YouTube video can be found at the following link: https://www.youtube.com/watch?v=pMMRE4Q2FGk

The lessons were intentionally chosen to cover very similar content in slightly different ways so that the same posttest could be used to test learning from both videos. The differences in production quality and style were not intentional but may have had a significant impact on the specific results of this study, which will be discussed below.

Outlines. Information from both video lessons was used to create two outlines, one consistent with the structure of the Crash Course lesson and one consistent with the structure of the Khan Academy lesson. The final outlines can be founds in Appendices G and H. The outlines were created by first outlining the two lessons independently and identifying information that was common to both lessons. Next, information that was unique to the Crash Course lesson was added to the Khan Academy outline, and information that was unique to the Khan Academy lesson was added to the Crash Course outline. The result was two outlines that contained identical information but that were structured differently according to the different video lessons.

Test of Prior Knowledge. The test of prior knowledge was designed to quickly gauge participants' familiarity with the topic of the lessons without creating a testing effect. Participants were asked to indicate how many out of 12 types of memory they felt they could define (these are listed in Table 5). After watching the lesson, participants also rated on a 1-5 scale how much of the information they knew before watching the video or studying the outlines (1=none of it, 5=all of it). A composite prior knowledge score was created by summing the number of memory types participants checked off with the self-rating. Scores ranged from 1-17, with an average score of 9.4 (*SD*=4.4) across all participants (median=9).

Before watching the video	*After watching the video*
Which of the following terms do you feel confident you could define?	How much did you know about this topic before watching the video?
○ Explicit memory	5 – a lot
○ Implicit memory	4
○ Semantic memory	3
○ Episodic memory	2
○ Declarative memory	1 – nothing
○ Procedural memory	
○ Long-term memory	
○ Short-term memory	
○ Working memory	
○ Sensory memory	
○ Iconic memory	
○ Echoic memory	

Table 5. Test of prior knowledge used in Experiment 8.

Retention test. The complete posttest (including both retention and transfer) can be found in Appendix I. The retention test was designed to assess what participants remembered from the lesson and consisted of 2 open response questions and 6 multiple-choice questions. The questions were written so that they were answerable after having watched either video. The first open response question asked participants to describe how information gets into long-term memory, using as much detail as possible. The second question asked participants to list as many different kinds of memory as they could remember and to describe what makes them different. The multiple choice questions focused on definitions of the types of memory that were presented in both lessons: sensory, short-term, long-term, explicit, implicit, episodic, and procedural. Open response and multiple-choice performance were analyzed separately, but also combined in a composite score (proportion correct) where each question was weighted equally.

Transfer test. The transfer test was designed to assess whether participants could identify types of memory used in different situations and apply that understanding to answer novel questions.

The test consisted of 4 open response questions and 6 multiple-choice questions. The open response questions were:

1. How is memory in the brain similar to memory in a computer? How is it different?
2. What could you do to improve your long-term memory?
3. You are driving to a new location and you've memorized the directions. Before you left, you studied a map and you also reviewed the turn-by-turn directions. How do you represent that information in your working memory?
4. Do you think all memories are stored in one spot in the brain in many spots? Why?

Answers to these questions were not presented directly in the lessons, but were hinted at. Two of the questions were hinted strongly in the Crash Course video, and two were hinted strongly in the Khan Academy video. Both outlines contained information that hinted at all 4 questions.

The multiple-choice transfer questions described hypothetical situations and asked participants to identify what kind of memory scenario described. These questions only tested types of memory that were described in both videos: sensory, episodic, procedural, short-term, and long-term. Open response and multiple-choice performance were analyzed separately, but also combined in a composite score (proportion correct) where each question was weighted equally.

Ratings. At the end of the lesson, participants rated several aspects of the lesson, including: how much they enjoyed the lesson, how much they knew about the content of the lesson before the study (included in the prior knowledge score), how much they thought they learned from the lesson, whether they would recommend the lesson to a friend, and whether they were interested in learning more about that topic. Participants also reported their age, gender, and the number of psychology courses they had ever taken.

Design & Procedure

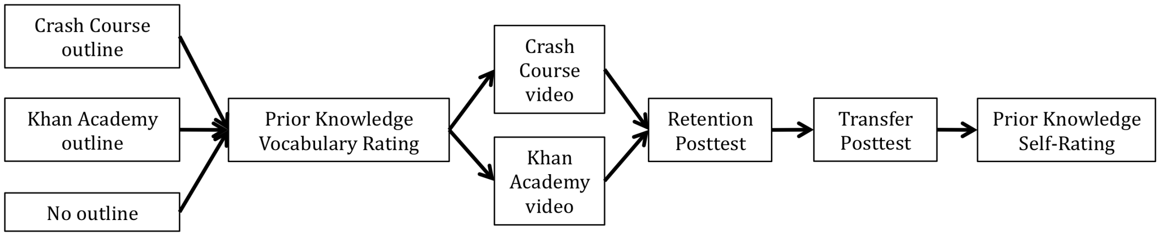


Figure 28. Design and procedure used in Experiment 8.

The design and procedure for Experiment 8 are shown in Figure 28. Participants were randomly assigned to one of six conditions created by crossing Outline (3: Crash Course, Khan Academy, or none) with Video (2: Crash Course or Khan Academy) for a 3x2 between-subjects design. For the purpose of some analyses, these 6 conditions were collapsed into 3 between-subjects conditions: *match* (consistent outline, *n*=30), *mismatch* (inconsistent outline, *n*=30), or *no outline* (*n*=26).

The *match* condition included subjects who saw an outline consistent with the video they watched (Crash Course outline and Crash Course video, *n*=15, or Khan Academy outline and Khan Academy video, *n*=15). The *mismatch* condition included subjects who studied an outline that was not consistent with the video they watched (Crash Course outline with Khan Academy video, *n*=15, or Khan Academy outline with Crash Course video, *n*=15). The *no outline* condition included subjects who did not study an outline before watching either the Crash Course video (*n*=13) or the Khan Academy video (*n*=13).

When participants arrived at the lab they were given a paper copy of the outline and asked to study it for the next 10 minutes because they would be tested on that information later. They were asked not to take notes and to use the whole 10 minutes to continuing reviewing the information. Participants in the two *no outline* conditions skipped this step and went straight to the video lesson. After studying the outline, participants complete the prior knowledge measure. All participants saw the same instructions:

> "In this video you will learn about different types of memory and the process by which we make new memories. At the end of the video, you will take a short test on this information. Before you begin, you will answer a few questions to help us gauge what you already know about memory."

Participants were then asked to indicate how many out of 12 different types of memory they felt confident they could define. After, they saw the following instructions before watching the video lesson:

> "Please treat this video like it is material you are trying to learn for a class. There will be a short test at the end to help us evaluate whether this lesson is a good study tool for this topic. Press play when you are ready to begin."

Immediately after watching the video, participants completed the posttest. Multiple-choice questions on the posttest (both retention and transfer) included feedback. This was included because pilot studies of the materials indicated that participants were more motivated to try hard on the posttest when they were given feedback. After the posttest, participants completed the ratings of the lesson, including the self-rating of how much information they knew before the study and an open response question asking whether they were taking or had ever taken a psychology course, and if so, which one(s). Participants also reported their age and gender.

Results

Overall retention scores were non-normally distributed ($W(86)=0.941$, $p=0.001$). The mean proportion correct on retention was 0.71 with a standard deviation of 0.17, but the distribution was negatively skewed. Transfer scores, however, were normally distributed ($W(86)=0.989$, $p=0.687$) with a mean of 0.58 and a standard deviation of 0.18.

Analysis of Match

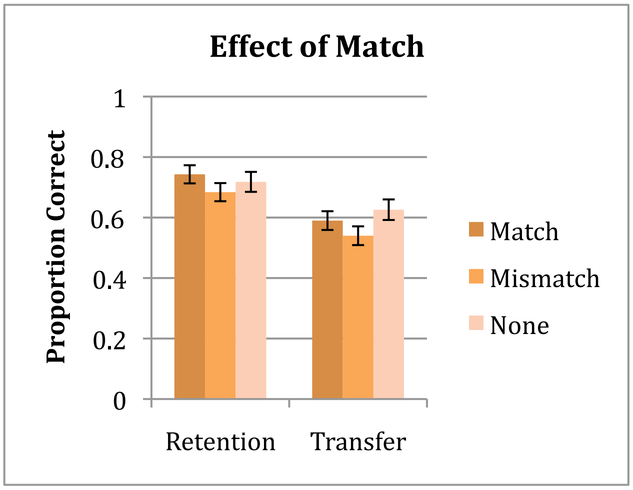


Figure 29. Mean proportion correct on the retention and transfer in Experiment 8 test by "match" category with prior knowledge as a covariate (evaluated at PK=9.37). Error bars indicate standard errors.

The effect of match (3: match, mismatch, no outline) on retention and transfer posttest scores was analyzed using a one-way MANCOVA with prior knowledge score as a covariate. Mean proportion correct on retention and transfer by match condition is show in Figure 29. Match had no significant effect on either retention (*F*(2,82)=1.01, *p*=0.368) or transfer (*F*(2,82)=1.74, *p*=0.181), but prior knowledge was marginally significant for retention (*F*(1,82)=3.76, *p*=0.056) and strongly significant for transfer (*F*(1,82)=13.00, *p*=0.001).

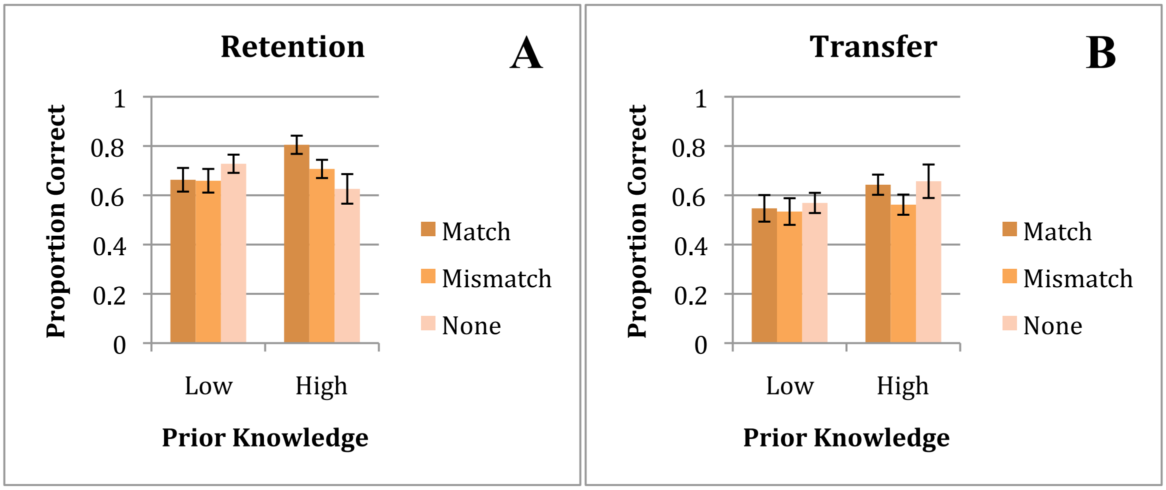


Figure 30. Mean proportion correct on retention (A) and transfer (B) by match condition and low or high prior knowledge in Experiment 8. Error bars indicate standard errors.

In order to test whether prior knowledge might interact with match condition, I used a median split to separate participants into high- and low-prior knowledge groups. Participants with a prior knowledge score of 9 or higher (*n*=45) were in the high-prior knowledge group, and participants with a score of 8 or lower (*n*=41) were in the low-prior knowledge group. I then used a two-way MANOVA with prior knowledge split (2: high, low) and match condition (3: match, mismatch, no outline) as between-subjects factors and retention and transfer as outcomes. The mean proportion correct in these groups can be seen in Figure 30. There was an interaction between match and prior knowledge split for retention (*F*(2,80)=3.45, *p*=0.037) but not for transfer (*F*(2,80)=0.290, *p*=0.749). The main effect of prior knowledge split was marginally significant for transfer (*F*(1,80)=2.90, *p*=0.093) but not for retention (*F*(1,80)=0.629, *p*=0.430). The main effect of match was not significant for either retention (*F*(2,80)=0.988, *p*=0.377) or transfer (*F*(2,80)=0.895, *p*=0.413).

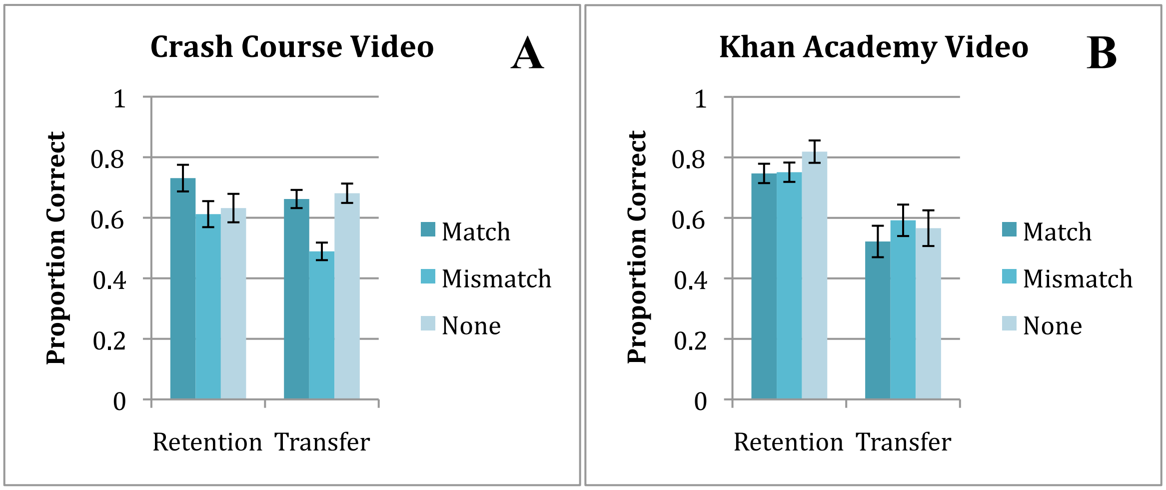


Figure 31. Mean proportion correct on the retention and transfer tests in Experiment 8 by "match" category for participants who watched the Crash Course video (A) and those who watched the Khan Academy video (B) with prior knowledge as a covariate (evaluated at PK=9.37). Error bars indicate standard errors.

In order to better approximate the original Mannes and Kintsch (1987) study, in which learning of one source (the article) was assessed but learning from the other source (the encyclopedia entry) was not, I conducted two additional analyses of the effect of match on retention and transfer – one for just the Crash Course video and one for just the Khan Academy video. The mean proportion correct for retention and transfer by match can be seen separately for the Crash Course and Khan Academy videos in Figure 31.

For the Crash Course video, the effect of match was statistically significant for transfer ($F(2,39)=12.57$, $p<0.001$), but not for retention ($F(2,39)=2.10$, $p=0.136$). The effect of prior knowledge was also significant for transfer ($F(1,39)=7.32$, $p=0.010$), but was only marginally significant for retention ($F(1,39)=3.92$, $p=0.055$).

For the Khan Academy video, the effect of match was non-significant for both retention ($F(2,39)=1.21$, $p=0.309$) and transfer ($F(2,39)=0.48$, $p=0.623$). The effect of prior knowledge, however, was significant for both retention ($F(1,39)=6.06$, $p=0.018$) and transfer ($F(1,39)=4.70$, $p=0.036$).

Analysis of Video x Outline

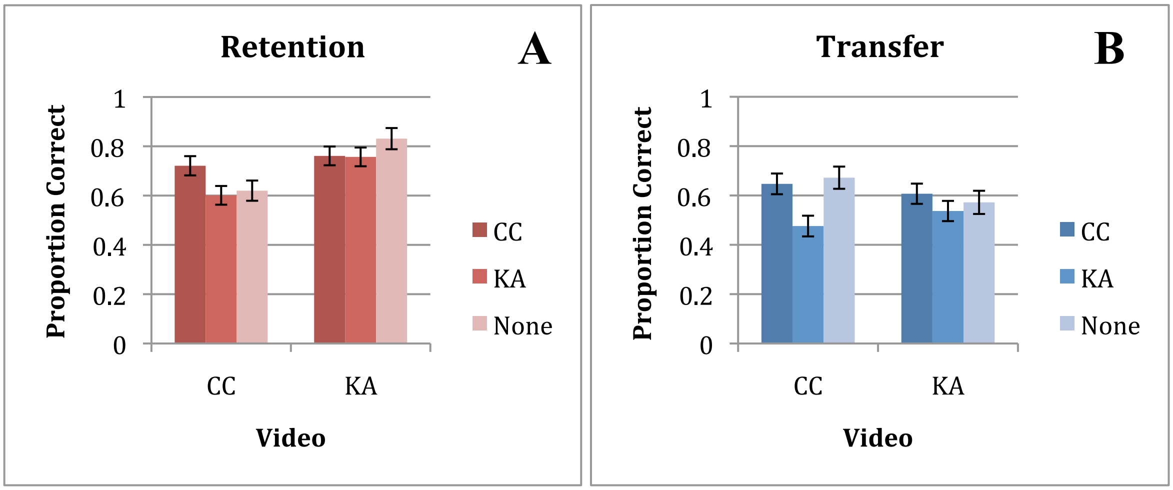


Figure 32. Mean proportion correct on the retention test (A) and transfer test (B) in Experiment 8 by Video and Outline with prior knowledge as a covariate (evaluated at PK=9.37). Error bars indicate standard errors.

The different effects of match for the Crash Course and Khan Academy video suggested that the specific videos and outlines might have a more significant effect on retention and transfer than the match condition. To test this hypothesis, I used a 2 (Video: Crash Course, Khan Academy) by 3 (Outline: Crash Course, Khan Academy, none) by 2 (Outcome: Retention, Transfer) MANCOVA with prior knowledge score as a covariate. Mean proportion correct on retention and transfer by video and outline can be seen in Figure 32.

For retention, the effects of prior knowledge ($F(1,79) = 9.45$, $p=0.003$) and video ($F(1,79) = 17.47$, $p<0.001$) were significant, but the effect of outline was not ($F(2,79) = 1.44$, $p=0.243$). There was a marginally significant interaction between outline and video ($F(1,79) = 2.53$, $p=0.086$). For participants who watched the Crash Course video, retention scores were higher when they studied the Crash Course outline ($M=0.73$, $SE=0.044$) than when they studied the Khan Academy outline ($M=0.61$, $SE=0.043$) or no outline at all ($M=0.63$, $SE=0.047$, $t(26)=2.09$). For participants who watched the Khan Academy video, retention scores were higher when they did not study an outline ($M=0.82$, $SE=0.037$) than when they studied either the Crash Course ($M=0.75$, $SE=0.032$) or the Khan Academy outline ($M=0.75$, $SE=0.032$).

For transfer, the main effects of prior knowledge and outline were significant (Prior Knowledge: $F(1,79) = 10.53$, $p=0.002$; Outline: $F(2,79) = 5.25$, $p=0.007$), but the effect of video was non-significant ($F(1,79) = 0.54$, $p=0.464$). The interaction between video and outline was not significant ($F(1,79) = 1.80$, $p=0.173$).

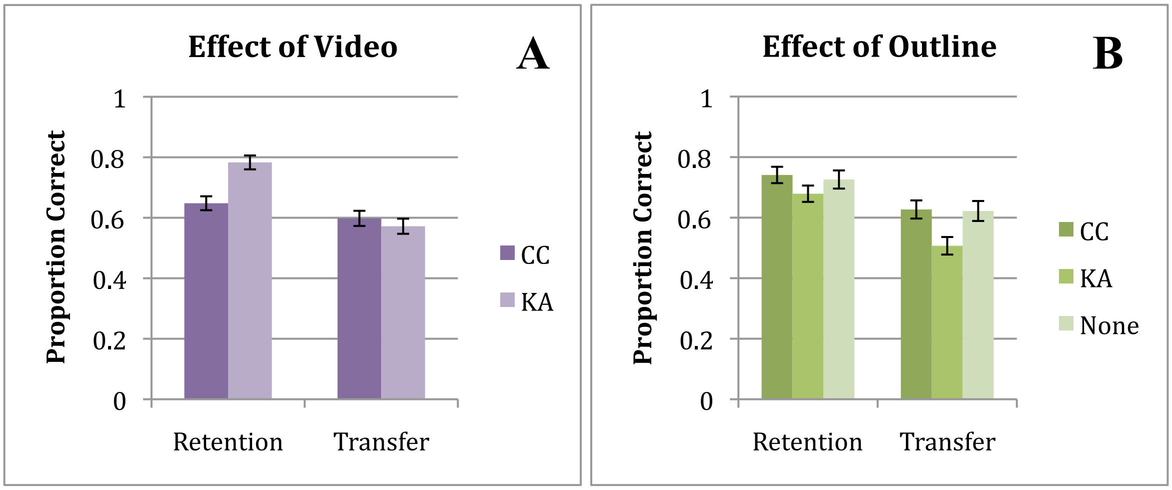


Figure 33. Mean proportion correct on the retention and transfer tests in Experiment 8 by Video (A) and by Outline (B) with prior knowledge as a covariate (evaluated at PK=9.37). Error bars indicate standard errors.

Overall there was a strong effect of Video on retention ($F(1,79) = 17.47$, $p<0.001$) and of Outline on transfer ($F(2,79) = 5.25$, $p=0.007$). Figure 33 shows performance collapsed across outlines (A) and videos (B) to highlight the effect of video on retention and outline on transfer. Retention scores were significantly higher for the Khan Academy video ($M=0.78$, $SE=0.023$) than for the Crash Course video ($M=0.65$, $SE=0.023$, $F(1,83)=15.75$ $p<0.001$). Transfer scores were significantly higher for the Crash Course outline ($M=0.63$, $SE=0.03$) and no outline ($M=0.63$, $SE=0.033$) conditions than for the Khan Academy outline ($M=0.51$, $SE=0.030$).

Discussion

The results of this study revealed no effect of outline consistency on either retention or transfer, but did indicate that the specific videos and outlines that participants studied significantly impacted their posttest performance. In general, retention was higher for participants who watched the Khan Academy video than for participants who watched the Crash Course video. This is not surprising, since the Khan Academy video is better designed to support retention of information. In the Khan Academy video information stayed on screen the entire time, and the video window was utilized well to organize information. In the Crash Course video, information disappeared from the screen shortly after it was presented, and there was no visual schematic that was used to organize all of the information in the lesson. In addition, many participants reported that the pace of the Crash Course lesson was a bit too fast to follow. Although both videos are designed to teach and review information, the Crash Course style video seems intentionally designed to be rewatched. There is no way that one could catch all of the information and jokes in the video from a single viewing.

The video participants watched did not have a significant effect on transfer performance, but the outlines they studied did significantly affect transfer. Participants tended to score higher if they had studied the Crash Course outline or no outline than if they had studied the Khan Academy outline. This is also not very surprising because the answers to the transfer questions were more salient in the Crash Course outline. Both outlines contained identical information, but the details that were relevant to the transfer questions tended to be presented as major headings in the Crash Course outline and as lower-level details in the Khan Academy outline. This also reflected the presentation of information in the Crash Course video, which suggests that the salience of the information (either in the video or the outline) is what contributed to the increased performance on those questions.

The results do not replicate the original Mannes and Kintsch (1987) findings, but they do show that advance organizers can influence learning. For the Crash Course video a matching outline increased retention but had no effect on transfer, and a mismatched outline had no effect on retention, but actually decreased transfer performance. For the Khan Academy video, neither type of outline affected either retention or transfer. It is somewhat surprising that neither video seemed to benefit from a mismatched outline for either retention or transfer. If one video better supported retention and the other outline better supported transfer, one would expect that the combination of the two (which would be a mismatched condition) would improve performance on both measures. The results of this study, however, showed that a mismatched outline *never* benefitted participants, and if it had any effect it was detrimental.

One possible explanation for this result is that mismatched outlines increased the perceived amount of material to be learned. When participants saw that the video they were watching was different from the outline they studied, they may have *felt* that they now had twice as much information to remember. The perceived increase in to-be-learned information may have increased anxiety and also lead participants to believe that all answers on the test were contained directly in the information presented. This explanation is supported by the fact that participants who studied an outline were more likely to write in their transfer question answers that this information had not be presented; participants who did not study an outline were much more likely to attempt to answer those questions.

Limitations

One limitation of this study was that participants spent relatively little time and effort studying the outlines. Although the length of the outlines and the lessons in this experiment were similar to the length of the outlines and article used in the original study, participants spent far more time studying the outlines in the original study. In Mannes and Kintsch's (1987) design, participants spent approximately 30 minutes reading the outlines and answering questions about them. Those activities probably helped ensure that participants had a solid understanding of the outline before moving on to the article. In this experiment, however, participants only spent 10 minutes studying the outline and were not asked to answer any questions about it or demonstrate any understanding of its contents. A closer replication of the original study should include similar tasks related to outline study before watching the video lessons.

Another probable limitation was the choice of topic for the lessons. In the original study, participants learned about uses of bacteria in industry, and the researchers were careful to select participants who did not have extensive experience in college-level biology. In this study, the topic of human memory was chosen because I was familiar enough with the material to feel confident that I could write an appropriate posttest, but my participants were somewhat familiar with the topic before the study. Participants were recruited through the psychology department subject pool, which is composed almost entirely of students who are currently enrolled in psychology courses. I collected data on the number of psychology courses the participants had ever taken, and the mean number of courses across all participants was 2.6 (SD=2.4). Although not every psychology course covers the topic of human memory, introductory courses definitely do, and introductory psychology was the most commonly reported course that participants had taken. This was not surprising, given that it is generally the first psychology course one takes, but it does indicate that nearly all participants had more familiarity with this topic than the average citizen does. As other experiments in this dissertation have shown, prior knowledge about a topic has a significant influence on subsequent learning, so it is possible that the lack of an effect of outline consistency was related to the participants' prior knowledge.

Future Directions

Future work on the effect of advance organizers on video learning should address several key concerns that were raised by this study: how (and how much) participants interact with the advance organizer, participants' prior knowledge about the lesson topic, and the specific features of the videos and tests that might influence performance. As discussed above, the way participants interact with an advance organizer is probably just as important as the content and format of that organizer. It does not matter how good the organizer is if students don't attend to it or understand it in a way that will support further learning.

The content of the lessons is also an important consideration for future studies. Advance organizers may be more useful when participants do not already have some knowledge that could be used to understand the lesson. By choosing topics with which students may be less familiar (or by choosing participants with varying degrees of familiarity with the topic) we may better understand what kinds of advance organizers are most useful when participants have no prior knowledge, some knowledge, or a lot of knowledge. There may also be important differences related to the type of content in the lesson – whether it is aimed at teaching a general concept, a step-by-step process, or something else.

Lastly, future studies of video learning should take into account the specific features of the videos themselves that might support or hinder learning. For example, the differences observed between Crash Course and Khan Academy – such as the speed of narration, how long information is displayed on screen, and the presence of an on-screen narrator – had effects on posttest performance that were stronger than the planned experimental manipulation. If our goal is to develop a better understanding of what advance organizers work when and for whom, these features of video lessons are sure to play a large role in understanding that interaction.

The results discussed here do not disprove Mannes and Kintsch’s findings or their interpretation that consistent organizers provide a schema while inconsistent organizers enhance prior knowledge. Instead, they simply suggest that there are many additional factors that contribute to learning that cannot be ignored. Understanding those factors will help us understand how to teach more effectively with a greater variety of materials and to a greater variety of students.

Chapter 6: Conclusions

This dissertation represents a theoretical and empirical exploration of learning from instructional video. The experiments described here contribute to the literature on multimedia learning both through empirical explorations of principles of multimedia instruction and through theoretical connections to the text comprehension literature. The goal of these studies was to conceptually replicate several findings from text comprehension and the cognitive theory of multimedia learning using video lessons. Although many effects failed to replicate in the experiments presented here, they still provide important insights into learning from instructional video.

The first important conclusion from this dissertation is that effects from text comprehension and CTML are not straightforwardly applicable to video lessons. Although many of the manipulations I explored in these experiments are robustly supported in their source literatures, they did not easily translate to the video lessons I chose. This could be for many reasons, but two strong possibilities are (1) that video is different in important ways from text and from narrated animations or (2) the interpretation of the original studies is oversimplified, or perhaps both.

With respect to the first possibility, there were several unforeseen features of videos that likely affected learning in ways that texts and animations do not. One such difference was the speed at which information is presented. Texts may differ in the coherence of the writing, but the reader determines the speed at which it is read. When watching a video, participants have no such control over the speed of narration. For example, the narrator in the Crash Course videos (used in all 8 experiments) speaks very quickly and the video is edited to eliminate almost all gaps in speech. This results in a video that moves quickly and does not allow time to think because viewers are just trying to keep up with the narration. It is difficult to assess manipulations of coherence when the problem students are experiencing has to do with speed, not explanation.

Speed is arguably an issue for all types of multimedia lessons that involve narration, but to varying degrees depending on the material and the length of the lesson. Mayer (2008) showed that segmenting a narrated animation can improve learning relative to a continuous lesson, but he tested animations where the segments were only 8-10 seconds long. In my experiments, longer segments did not seem to have the same benefit for learning. It is likely that the students' experience of trying to comprehend a longer, faster, more complicated lesson was affected by more than just pauses and headings.

Another way that video differs from other learning materials is in the visual organization of information. There are a limited number of ways that text can be organized on a page, but the visual component of a video can take many forms and does not always support learning. This was most clearly illustrated in the differences between the Crash Course and Khan Academy videos used in Experiment 8. Although the content covered was similar, the visual presentation was very different. The Khan Academy video used the visual component of the lesson to display information continuously and to present key concepts in a planned spatial organization. The Crash Course video, on the other hand, frequently switched between animations, text slides, and the on-screen narrator. Although it is useful to know what kinds of videos might better support learning than others, differences like these make it difficult to apply findings from a single study to many kinds of materials. The characteristics of the video itself matter as much as (if not more than) the specific manipulation being tested.

The second reason that findings from text comprehension and CTML may not translate directly to video learning is related to the way the original findings are interpreted. Specifically, the general nature of many interpretations could lead to replication attempts that miss the important features of the original study. For example, Mannes and Kintsch (1987) interpreted their findings to suggest that consistent organizers benefit retention while inconsistent organizers benefit transfer. In the context of their study, it was true that participants who saw the outline consistent with the text performed better on retention than participants who studied the inconsistent outline, and also true that participants who studied the inconsistent outline performed better on transfer questions than participants who studied the consistent outline. However, those results were the product of studying particular materials and being asked particular questions on the posttest – not simply the differences in consistency. It is entirely possible that slightly different materials or slightly different questions might have yielded different results.

This is, of course, true of any experiment. But failing to acknowledge the influence of specific materials in the outcome of the study can lead to overextension and over-application of a finding. According to Google Scholar, the Mannes and Kintsch (1987) study has been cited 283 times, but I have not found a single replication of their study. I used the idea that consistent organizers improve retention and inconsistent organizers improve transfer to design the materials for my study, but was not able to replicate the effect either. This does not mean that such an effect does not exist, but suggests that it is dependent on certain kinds of materials and questions. The fact that no replication of the original study has been published suggests that perhaps the effect was not about consistency at all, but had to do with other aspects of the design, materials, or procedure.

Mayer's (2008) review of work supporting his principles of multimedia instruction suffers from similar problems. Although Mayer identifies multiple studies to support the principles in his theory, those studies use very similar types of materials and measures. As a result, the replication of the segmenting effect or the signaling effect is a replication of the effect *with those particular materials*. Again, this does not mean that the effect does not exist! It only suggests that much more work is needed to understand how the principles of multimedia instruction apply to other (very different) types of multimedia. The fact that segmenting and signaling improve learning with some materials does not mean that the benefits of segmenting and signaling will be detectable with *all* materials.

Limitations

These limitations to the interpretation of the source literatures are limitations for the experiments in this dissertation as well. The importance of appropriate material was a point of difficulty for many reasons. First, choosing an appropriate topic depended a lot on who the participants would be. When recruiting participants through Mechanical Turk, I had few concerns that the topic I might pick would be too easy, but did wonder if some topics might be too difficult or not engaging enough. Several Mturk participants contacted me to say how much they enjoyed the Chemical Mind video, but others complained about the difficulty of following the excretory system lesson. When recruiting UCLA students, however, I had the problem of finding material that was novel enough and difficult enough to produce a range of scores on the posttest. It is likely that I aimed too low with the lessons on human memory, given that many students had already studied this material before.

An additional problem with choosing the right lesson topic is creating an appropriate posttest. Although many instructional videos exist online, few (if any) are associated with specific tests to assess viewers' learning. This meant that I had to write a new posttest for every topic I chose to use in the study. Writing posttests is not simple and requires much iteration to find a set of questions that produces normally distributed scores and captures what students actually learn. My ability to do this well was further limited by my own knowledge of the topics. My familiarity with the topics used in Chapters 3 and 5 (neuron communication and human memory, respectively) made this easier, but the video about kidney function required an additional investment of time spent understanding the lesson content well and then writing good questions. Without the aid of an expert in the field, writing good posttests is a big investment. It is not surprising, then, that many researchers stick to the same topic once they have found a measure that works.

Future Directions

In spite of the difficulties finding significant effects, there are many avenues for future research that can further build our understanding of learning from instructional video. One of these avenues is the use of questions both as signals within a lesson and as advance organizers. As reviewed in the introduction, there is substantial evidence from text comprehension that questions can be used to help students learn from text more effectively. Investigating the use of questions in video lessons could also help clarify how principles of CTML might help foster generative processing.

Whatever specific questions ones pursues related to improving learning from video, another set of questions should be asked in parallel regarding the materials and measures being used. These questions should include: how can we better measure and/or control participants' prior knowledge about a topic? How can we better measure and/or control the coherence of a video lesson? What characteristics of instructional video are likely to help or hinder learning, and for whom? Learning more about the influence of specific features of videos on learning should help us better understand and apply findings to different kinds of video lessons. By gaining a better understanding of how students learn from instructional video and how teachers can enhance video to improve what students learn, I hope to advance both basic research on cognition and learning as well as the applied science of instruction.

Appendix A

Transcript of lesson used in Experiments 1 and 2.
Text from informative pauses is italicized.
The original video can be found on YouTube at:
https://www.youtube.com/watch?v=W4N-7AlzK7s

Say it's late at night, you're home alone drifting off to sleep, just, entering that dream about Fritos, and then suddenly there's a banging at the door! Suddenly you're wide awake and it feels like your heart's gonna explode. You jump up ready to run out the back door, possibly grab a Phillips head screwdriver and stab it into the darkness until it sticks into something.

Now whether it's a Weeping Angel or your neighbor looking to borrow a can of beans, it doesn't really matter because when you heard that sudden noise, your startled brain released an icy typhoon of chemicals. And everything that's now going through your mind, like your urge to flee, your urge to defend yourself, that internal debate about whether Weeping Angels are even real and "Woah! Where's the cat?" All that? Is just a result of those chemicals.

Our brains and our nervous systems and the substances they produce and are always bathed in are amazingly complex nuanced systems. And even though we're always talking about our mental activities being somehow separate from all the biological stuff going on in our bodies, in reality, the moods, ideas, impulses, that flash through our minds are spurred by our biological condition. As psychologists like to say, "Everything psychological is biological."

[1] "Everything psychological is biological" means that all mental processes are based in biological processes.

So one way to understand how your mind works is to look at how the chemistry of your body influences how you think, sense, and feel about the world around you. To do that, we begin at the simplest level, the system with the smallest parts, it's all about the neuron, baby. Neurons, or nerve cells, are the building blocks that comprise our nervous systems. Neurons share the same basic makeup as our other cells, but they have electrochemical mojo that lets them transmit messages to each other.

[2a] What he means by "electrochemical mojo" is that neuron communication happens in two parts: an electrical part and a chemical part.
[2b] Both parts are carried about by chemicals, but the first part depends on the electrical charge of ions inside the cell, and the second part depends on the specific types of chemicals that jump across to neighboring neurons.

Your brain alone is made up of billions of neurons, and to understand why we think or dream or do anything, you gotta first understand how these little transmitters work. You actually have several different types of neurons in your body, from ones that are less than a millimeter long in your brain to ones that run the whole length of your leg! Yes, you have cells as long as your legs, which is nothing compared to the hundred and fifty feet the nerve cells of some dinosaurs had to be, I'm getting off topic, sorry.

No matter how big a nerve is, they all have the same three basic parts: the soma, dendrites, and axon. The soma, or cell body, is basically the neuron's life support; it contains all that necessary cell action like the nucleus, DNA, mitochondria, ribosomes, and such. So, if the soma dies, the whole neuron goes with it. The dendrites, as bushy and branch-like as the trees they're named after, receive messages and gossip from other cells. They're the listeners, whispering what they hear back to the soma. The axon is the talker. This long, cable-like extension transmits electrical impulses from the cell body out to other neurons or glands or muscles.

[3] The electrical impulse he just mentioned is the electrical part of the electrochemical signal that neurons use to communicate.

Whereas the dendrites are short and bushy, the axon fiber is long, and, depending on what type of neuron it is, is sometimes encased in a protective layer of fatty tissue, called the myelin sheath. It's almost like an insulated electrical wire, the myelin sheath speeds up the transmission of messages, and if it degrades, as it does with those affected with multiple sclerosis, those signals are degraded as well, eventually leading to lack of muscle control.

[4] Review: The dendrites "listen" for messages from other neurons, the soma keeps the neuron alive, and the axon "talks" to other neurons.

Neurons transmit signals either when stimulated by sensory input or triggered by neighboring neurons. The dendrites pick up the signal and activate the neuron's action potential, or firing impulse, that shoots an electrical charge down the axon to its terminals and towards the neighboring neurons.

[5a] The "action potential" is the name for the electrical impulse that forms the electrical part of the electrochemical signal. On this diagram it is labeled as the "neural impulse".
[5b] Action potentials always travel down the axon away from the cell body. Whether or not a neuron fires an action potential depends on the messages it receives from other neurons.

The contact points between neurons are called synapses. All those bushy little dendrites are decorated with synapses that almost but don't quite touch the neighboring axon in the tiniest game of "I'm not touching you!" of all time. They're less than a millionth of an inch apart. And that microscopic cleft is called the synaptic gap.

*[6] It's important to note that the synapse is not actually *part* of the neuron - it's the name for the gap between neurons.*

So, when an action potential runs down to the end of an axon, it activates the chemical messengers that jump that tiny synaptic gap, flying like that little air kiss and landing on the receptor sites of the receiving neuron. Those messengers are neurotransmitters.

Although neurotransmitters slide right into their intended receptors like a key into a lock, they don't stay bonded to the receiving neuron. They just sort of pop out, having excited or inhibited the receiving neuron's trigger, then the extras immediately get reabsorbed by the neuron that released them in the first place in a process called reuptake.

[7a] The release and reuptake of neurotransmitters forms the chemical part of the electrochemical signal.
*[7b] The electrical part depends on the ions *inside* the neuron, but the chemical part depends on the neurotransmitters (NTs) that are released into the synaptic gap. Different NTs cause different electrical changes in the receiving neurons.*

Kinda like, "Here you go, oh, psych!"

So neurons communicate with neurotransmitters which in turn cause motion and emotion; they help us move around, make jazz hands, learn, feel, remember, stay alert, get sleepy, and pretty much do everything we do. Some of them just make you feel good, like the endorphins we get flooded with after running ten miles or falling in love or eating a really good piece of pie. We've got over 100 different kinds of these brilliant neurotransmitters -- some are excitatory and others are inhibitory, and all are good reminders that everything psychological is also biological.

[8] Review: Neurotransmitters are chemicals that jump back and forth across the synaptic gap in order to carry messages from one neuron to the next.

Excitatory neurotransmitters rev up the neuron, increasing the chances it will fire off an action potential.

[9] Excitatory NTs make the receiving neuron more likely to fire an action potential. They increase the chance that the message will be passed on.

Norepinephrine is one you're probably familiar with, it helps control alertness and arousal. Glutamate is another, involved in memory, but an over-supply of it can wig out the brain and cause seizures and migraines which is why some people are sensitive to all that MSG, or monosodium glutamate, in their Ramen.

Inhibitory neurotransmitters on the other hand, chill neurons out, decreasing the likelihood that the neuron will jump into action.

[10] Inhibitory NTs make the receiving neuron less likely to fire an action potential. They decrease the chance that the message will be passed on.

GABA– gamma-aminobutyric acid– is a major inhibitory neurotransmitter, and you've probably heard of serotonin which affects your mood and hunger and sleep. Low amounts of serotonin are linked to depression, and a certain class of antidepressants help raise serotonin levels in the brain. Some neurotransmitters like acetylcholine and dopamine play both sides and can both excite or inhibit neurons depending on what type of receptors they encounter. Acetylcholine enables muscle action and influences learning and memory; Alzheimer's patients experience a deterioration of their acetylcholine producing neurons. Dopamine, meanwhile, is associated with learning, movement, and pleasurable emotions, and excessive amounts of it are linked to schizophrenia as well as addictive and impulsive behavior.

[11a] Some examples of excitatory NTs were norepinephrine and glutamate. Some examples of inhibitory NTs were GABA and serotonin.
[11b] Some NTs, such as acetylcholine (ACh) and dopamine, can be both excitatory and inhibitory, depending on the situation.

So neurotransmitters are basically your nervous system's couriers. But they aren't the only chemical messengers delivering the news; they've got some competition brewing in the endocrine system. And if you've been through puberty, you know what I'm talking about: hormones.

Like neurotransmitters, hormones act on the brain, and indeed some of them are chemically identical to certain neurotransmitters. Hormones affect our moods, arousal, and circadian rhythm, they regulate our metabolism, monitor our immune system, signal growth, and help with sexual reproduction. You could say that most of them boil down to the basics: attraction, appetite, and aggression.

Whereas neurons and synapses flick on and off, sending messages with amazing speed, the endocrine system likes to take its time, delivering the body's slow chemical communications through a set of glands that secrete hormones into the bloodstream where they're ferried to other tissues, especially the brain.

[12a] Review: Hormones are secreted into the bloodstream, where they travel throughout the body, including to the brain.
[12b] Hormones can affect lots of things, but in general you could summarize their effects into feelings of attraction, appetite, and aggression.
[12c] Some hormones are almost identical to neurotransmitters but they are different in important ways.

So while the nervous and endocrine systems are similar, in that they both produce chemicals destined to hit up certain receptors, they operate at very different speeds. It's like, if the nervous system wants to get in touch with you, it sends you a text. But if the endocrine system has a message, it will like lick the stamp, and put it on, and write your address, and then a note and a pen on paper, and then fold it up and put and mail it to you with the Post Office. But fast isn't always better, and your body will remember that letter longer than the text. Hormones, they linger. Which helps explain why it takes some time to simmer down after a moment of severe fright or anger.

[13] Review: While NTs are stored in the brain, communicate quickly, and have a short effect, hormones are secreted from other glands, communicate slowly, and have a longer effect.

And our endocrine systems have a few important hormone brewing glands. We've got a pair of adrenal glands snuggled up against our kidneys that secrete adrenaline, that famous fight or flight hormone that jacks up your heart rate, blood pressure and blood sugar, giving you that tidal wave of energy preparing you to run like heck or punch that charging baboon in the throat; the pancreas sits right next to the adrenal gland and oozes insulin and glucagon hormones that monitor how you absorb sugar, your bodies main source of fuel. Your thyroid and parathyroid glands at the base of your throat secrete hormones that regulate your metabolism and monitor your body's calcium levels; if you have testicles, they're secreting your sex hormones like estrogen and testosterone, and if you've got ovaries, they're doing that job.

And all those glands are super important, but there is one gland that rules them all, and in the darkness binds them: the pituitary gland. Although it's just a little pea-sized nugget hidden deep in the bunker of the brain, it is the most influential gland in this system. It releases a vital growth hormone that spurs physical development and that love hormone, oxytocin, that promotes warm, fuzzy feelings of trust and social bonding.

[14a] Review: The endocrine glands are the adrenal gland, the pancreas, the thyroid and parathyroid, the testicles or ovaries, and the pituitary gland.
[14b] The hormones Hank mentioned were adrenaline, insulin, glucagon, testosterone, estrogen, growth hormone, and oxytocin.

What really makes the pituitary the master gland is that its secretions boss around the other endocrine glands, but even the pituitary has a master in the hypothalamus region of the brain, which we will talk more about next episode.

*[15] It's important to note that the hypothalamus is NOT an endocrine gland. It is a part of the brain that *communicates* with the endocrine system by influencing the pituitary gland.*

So, AHHHHHHHHH! If I managed to scare you, sorry, but I'm illustrating a point. You have no control over being scared, but maybe now you do understand a little more clearly how your nervous and endocrine systems worked together to call the shots. First, the sensory input from your eyes and ears went to your brain, the simplest bits of your hypothalamus without even letting you analyze it and were like ahhhh, and then, that ran down the chain of command from your pituitary to your adrenal glands, to the hormone adrenaline, to the rest of your body and then back to your brain, which then realized that I was just messing with you and told everybody to just calm down for once!

The whole deal is a feedback loop: your nervous system directs your endocrine system which directs your nervous system, brain, gland, hormone, brain.

[16a] This is important! The central nervous system and the endocrine system influence and communicate with each other.
[16b] Sensory input travels to the hypothalamus by electrochemical signals (via neurons). The hypothalamus then stimulates the pituitary gland.
[16c] The pituitary gland releases hormones that communicate with other glands in the endocrine system (in this example, it's the adrenal gland specifically).
[16d] Those glands then secrete other hormones (in this example, adrenaline) which travel back to the brain via the bloodstream and provide new input to the neurons.

And of course each of these systems is fantastically complex. Way more than we can get into here. So, in our next lesson, we're gonna get all up in your brain, and delve deeper into the different components of your nervous system, find out what your old brain is, and learn about how much of your brain you actually use.

Appendix B

Posttest for Experiment 1

1. What are the major parts of the neuron? Check all that apply:
    a. Soma (cell body)
    b. Dendrites
    c. Axon
    d. Action potential
    e. Synapse

2. How do neurons communicate? Describe the steps in as much detail as you can:

3. Which of these is a neurotransmitter?
    a. Insulin
    b. Testosterone
    c. Oxytocin
    d. GABA

4. What is the difference between excitatory and inhibitory neurotransmitters?
    a. Excitatory NTs make the action potential move faster and inhibitory NTs make the action potential move slower
    b. Excitatory NTs make an action potential more likely and inhibitory NTs make an action potential less likely
    c. Excitatory NTs make you more alert and inhibitory NTs make you less alert
    d. Excitatory NTs make you think faster and inhibitory NTs make you think slower

5. Which of these is a hormone?
    a. Dopamine
    b. Serotonin
    c. Adrenaline
    d. Acetylcholine

6. What is the difference between neurotransmitters and hormones? Check all that apply:
    a. Neurotransmitters only travel in the synapses between neurons, but hormones travel throughout the body via the bloodstream
    b. Neurotransmitters have a very fast but short-lived effect; hormones are slower to affect the brain but their effects last longer
    c. Neurotransmitters are stored and released from neurons, but hormones are stored and released in specific endocrine glands
    d. Both hormones and NTs affect the nervous system, but only hormones can affect other tissues in the body

7. Which part of the brain directly influences the endocrine system?
    a. Hypothalamus
    b. Parathyroid
    c. Adrenal
    d. Pituitary

8. What does it mean to say that “everything psychological is biological”?

Appendix C

Transcript of the lesson used in Experiments 3-7
Timestamps and titles indicate locations where heading slides were inserted.
Timestamps are from the original YouTube video, which can be found here:
https://www.youtube.com/watch?v=WtrYotjYvtU

0:00 Segment 1: Homeostasis and Osmoregulation

One of the coolest and most important things that our bodies do is maintain this thing called homeostasis, the regulation of a stable internal environment, no matter where we are or what we're doing. After all, we put our bodies through a lot every single day: We're always adding food and liquid and chemicals, and we're constantly changing temperature and our levels of activity, but our bodies can roll with it. It's like, no big deal for them.

All of our organ systems have some hand in maintaining homeostasis. I mean, it's basically the thing that makes us not dead. But the excretory system (aka the urinary system), which includes the kidneys, the ureters, the bladder, and the urethra, is the star quarterback of the homeostasis team. That's because your excretory system is responsible for maintaining the right levels of water and dissolved substances in your body. This is called osmoregulation, and it's how our bodies get rid of the stuff we don't need (like the byproducts of metabolizing food), while also making sure we don't get dehydrated. It's the body's greatest balancing act, and your body is doing it right now, and all the time, as long as you're not dead.

1:08 Segment 2: Urea and Uric Acid

As with other organ systems we've talked about, not all excretory systems in the animal kingdom are created equal. Different animals excrete waste different ways based on their evolutionary history, what environments they live in, and what their hobbies and interests are. These factors all influence how an animal regulates water, and most metabolic waste needs to be dissolved in water in order to be excreted. The problem is, a main byproduct of metabolizing food is ammonia, which comes from breaking down proteins, and it's pretty toxic. So, depending on how much water is available to an animal and how easy it is for the animal to lug a bunch of water around inside it, animals convert this ammonia into either urea or uric acid.

Mammals like us, as well as amphibians, and some marine animals like sharks and sea turtles, convert ammonia into urea, a compound made from combining ammonia and carbon dioxide, in their livers. The advantage of urea is its very low toxicity. It can hang out in your circulatory systems for a while with no ill effects. But you have to have some extra water available to dissolve it and get rid of it. This isn't such a tall order, really, I mean peeing isn't a huge inconvenience, I mean, is it? It's not for me anyways.

Well, it would be, though, if you were a bird or an insect or a lizard living in the desert. Animals that have to be light enough to fly or don't have a bunch of spare water hanging around, convert ammonia into uric acid, which can be excreted as a kind of paste, so not a lot of water is needed. You've seen bird poop. If you haven't taken a close look, next time, do that. Just look. The white stuff in the bird droppings is actually the uric acid-y pee and the brown stuff is the poop. So, now that we've established what is and what is not bird poop, let's get down to the brass tacks of how humans get all of this urea out of our blood and into our toilets.

2:50 Segment 3: Kidney & Nephron

The excretory system starts with the kidneys, the organs that do all the heavy lifting, from maintaining those levels of water and dissolved materials in our bodies to controlling our blood pressure. And even though they do an amazing job – I'm not bad-mouthing your kidneys here – the way that they do it is frankly a little bit janky and inefficient. They start out by filtering out a bunch of fluid and the stuff dissolved in the fluid out of your blood, and then they basically re-absorb 99% of it back before sending that 1% on its way in the form of urine. Seriously, 99% gets re-absorbed.

On an average day, your kidneys filter out about 180 liters of fluid from your blood, but only 1.5 liters of that ends up getting peed out. So most of your excretory system isn't dedicated to excreting it's dedicated to re-absorbing. But the system works, obviously, I'm still alive. So we can't argue with that. Now it is time to get into the nitty gritty details of how your kidneys do all this, and it's pretty cool. But there's lots of weird words. So get ready.

Your kidneys do all this work using a network of tiny filtering structures called nephrons. Each one of your mango-sized kidneys has about a million of them. If you were – don't do this – but if you were to unravel all of your nephrons and put them end to end, they would stretch over 80 kilometers. This is where all the crazy action happens, so to understand how they work, we're just going to follow the flow, from your heart to the toilet.

4:08 Segment 4: Glomerulus & Bowman's Capsule

Blood from the heart enters the kidneys through renal arteries, and just so you know, whenever you hear the word "renal" it means you're dealing with kidney stuff. As the blood enters, it's forced into a system of tiny capillaries until it enters a tangle of porous capillaries called the glomerulus. This is the starting point for a single nephron.

The pressure in the glomerulus is high enough that it squeezes some of the fluid out of the blood, about 20% of it, and into a cup-like sac called the Bowman's capsule. The stuff that's squeezed out is no longer blood, it is now called filtrate. It's made up of water, urea, some smaller ions and molecules like sodium, glucose and amino acids. The bigger stuff in your blood, like the red blood cells and the larger proteins, they don't get filtered.

4:49 Segment 5: Proximal Convoluted Tubule

Now the filtrate is ready to be processed. From the Bowman's capsule, it flows into a twisted tube called the proximal convoluted tubule, which means "the tube near the beginning and that is all wind-y." WHY ARE WE SO BAD AT NAMING THINGS?! Anyways, this is the first of two convoluted tubules in the nephron. And these, along with other tubules we're talking about, are where the osmoregulation takes place. With all kinds of tricked out, specialized pumps and other kinds of active and passive transport, they re-absorb water and dissolved materials to create whatever balance your body needs at the time.

In the proximal tubule, it's mainly organic solutes in the filtrate that are reabsorbed like glucose, and amino acids, and other important stuff that you want to hang on to. But it also helps to re-capture some sodium, potassium and water we're going to want later.

5:34 Segment 6: Renal Cortex and Renal Medulla

From here, the filtrate enters the Loop of Henle, which is a long, hairpin-shaped tubule that passes through the two main layers of the kidney. The outermost layer is the renal cortex, that's where the glomerulus, bowman's capsule, and both convoluted tubules are, and the layer beneath that is the renal medulla, which is the center of the kidney. "Cortex," by the way, is Latin for tree bark, so whenever you see it in biology, you know that it's the outside of something. "Medulla," on the other hand, means “marrow” or “pith”, so you know that it's the inside. Just to help you remember this stuff.

[cut out the section from 6:05 to 7:36]

7:36 Segment 7: Loop of Henle

Alright, so, review time. We've squeezed some filtrate out of the blood, and re-absorbed some of the important organic molecules we want to keep. But most of the re-absorption action happens here, in the Loop of Henle, which does three really important things. One, it extracts most of the water that we need from the filtrate as it travels down to the medulla. Two, it pumps out the salts that we want to keep on the way back up to the cortex. And three, in the process of doing all that, it makes the medulla hypertonic, or super salty relative to the filtrate, creating a concentration gradient that will allow the medulla to draw out even more water one last time from the filtrate, before the final journey to the toilet begins. It's complicated and, again, kinda janky, but it's what allows us mammals to create urine that's as concentrated as necessary, using only the amount of water that our bodies can spare at the time.

8:21 Segment 8: Descending Limb

So first, filtrate starts going down the loop, and the thing to know here is that the membrane is highly permeable to water, not so much to salt or anything else, mainly water. Now, compared to the filtrate, the tissue of the medulla is already pretty salty. And as the filtrate processes, the surrounding tissue becomes increasingly hypertonic the farther down you go, the saltier it gets. So, applying everything we've learned about osmosis, you know that as the filtrate moves along, it loses more and more water through the membrane. By the time the filtrate gets to the bottom of the Loop, it's highly concentrated.

8:51 Segment 9: Ascending Limb

Now the filtrate enters the ascending end of the Loop, and here it's basically the same but in reverse. The membrane is NOT permeable to water, and instead it's lined with channels that transport ions like sodium, potassium and chlorine. And because the filtrate is so concentrated now, it's actually hypertonic compared to the fluid outside in the medulla. So as it ascends, huge amounts of salts start flowing out of the filtrate, which makes the renal medulla really, really, really salty. This salty medulla also creates a concentration gradient between the medulla and the filtrate, which we're going to need in the final step of pee-making.

9:23 Segment 10: Distal Convoluted Tubule

But first! Once the filtrate is back up in the cortex and out of the loop, it enters the second of our convoluted tubules, called the distal convoluted tubule, or "farther-away curly tube." While the first tubule worked mostly on reabsorbing the organic compounds in the filtrate, here the focus is on regulating levels of potassium, sodium, and calcium. This work is mainly done by pumps and hormones that regulate the reabsorption process.

9:47 Segment 11: Collecting Duct

By the time it's done, we've finally taken everything we want to keep out of the filtrate, so now it's mainly just excess water, urea and other metabolic waste. This stuff all gets dumped into collecting ducts that channel it back down to the center of the kidney, the medulla. And remember, the medulla is super-salty, right? Now more hormones kick in that tell the collecting ducts how porous to make their membranes. If the membranes are made very porous, more water is absorbed into the medulla, which makes the urine yes, we can start calling it urine now even more concentrated.

And here's a fun fact: If you've ever had one drink too many, you might've noticed that you start to pee a lot, and your pee is clear. That's because alcohol interferes with these hormones especially one called anti-diuretic hormone which tells the collecting ducts to be very porous so that you reabsorb most of the water. With those hormones all confused and out of commission, you just starting peeing out all kinds of water, which also means you're getting dehydrated, which means you're officially on a one-way trip to Hangover City. So, now you know why that happens.

10:45 Segment 12: Ureters, Bladder, & Urethra

Now at this point, the urine leaves both kidneys and flows down to the urinary bladder by tubes called ureters. Once in the bladder, the urine just sits around, waiting for us to decide when it's time to find a bathroom. And when that time comes, a little sphincter muscle relaxes and releases the urine from the bladder into a tube called the urethra, which empties out wherever you point it. So that's how your excretory system works!

And that's basically how it works for most mammals, although some modifications are made based on, again, where they live and what they do. For instance, kangaroo rats, which are tiny and adorable and live in the desert, have the most concentrated urine of any animal anywhere, because it can't spare the water. So it has a very, very long Loop of Henle that reabsorbs most of the water from the filtrate. On the other end of the spectrum, we have the beavers, who have very short Loops of Henle, because they're like, "Water reabsorption, schmater reabschmorption. Do you see what I do all day?" And so now you know the true origins of pee.

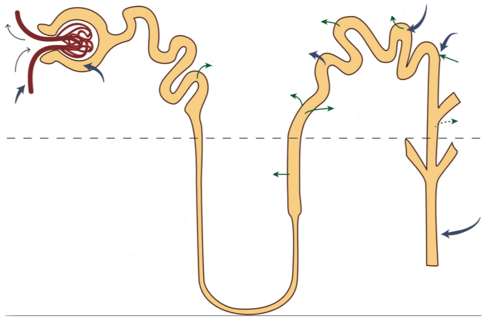
Appendix D

Posttest for Experiment 3

[Retention Questions]

1. How does the kidney filter blood? Describe the steps in as much detail as you can. Please include as many of the following words in your answer as you can:
   - concentration gradient,
   - osmosis,
   - water,
   - glucose,
   - ions,
   - urea,
   - filtrate,
   - glomerulus,
   - bowman's capsule,
   - loop of henle,
   - collecting duct,
   - proximal & distal convoluted tubes,
   - descending & ascending limbs

2. Name the parts of the nephron and describe the function of each part. You may use the image on the right to help you.



3. The glomerulus filters the blood from the ______ and the resulting filtrate collects in the _______ before it flows through the rest of the nephron.
   a. renal vein, proximal convoluted tubule
   b. renal vein, Bowman's capsule
   c. renal artery, proximal convoluted tubule
   d. renal artery, Bowman's capsule

4. What is the substance that travels through the nephron?
   a. Blood
   b. Ions
   c. Filtrate
   d. Water

5. Most of the water reabsorption happens in the _____.
    a. Proximal convoluted tubule
    b. Loop of Henle
    c. Distal convoluted tubule
    d. Collecting duct

6. The descending limb of the Loop of Henle is permeable to ____, while the ascending limb is permeable to ____.
    a. Filtrate, sodium
    b. Sodium, filtrate
    c. ions, water
    d. water, ions

7. While the descending limb relies on _______ to transport water, the ascending limb relies on _____ to transport ions.
    a. Osmosis; active transport
    b. Active transport; osmosis
    c. Osmosis; diffusion
    d. Active transport; diffusion

8. Which one of these is NOT a difference between the proximal and distal convoluted tubules?
    a. the proximal convoluted tubule reabsorbs glucose and amino acids while the distal convoluted tubules does not
    b. the proximal convoluted tubule reabsorbs more nutrients and water than the distal convoluted tubule
    c. the proximal convoluted tubule helps regulate water (osmoregulation) but the distal convoluted tubule does not
    d. the proximal convoluted tubule is closer to the glomerulus than the distal convoluted tubule

[Transfer Questions]

9. What would happen if we make the collecting duct's membrane more porous?
    a. ions from the medulla will leak back into the collecting duct and the urine will be more concentrated
    b. more water from the filtrate will be reabsorbed into the medulla and the urine will be more concentrated
    c. both ions and water from the medulla will leak back into the collecting duct and the urine's color will not change but there will be more urine
    d. both ions and water will be reabsorbed into the medulla and the urine's color will not change but there will be less urine

10. What would happen to the amount of water and ions reabsorbed by the nephron if someone did not have a loop of Henle?
    a. the person will not feel a difference because the amount of water and ions reabsorbed in the loop of Henle is nearly negligible
    b. the person will not feel a difference because the proximal convoluted tubule, distal convoluted tubule and collecting duct will provide enough reabsorption of water and ions
    c. the person will lose a lot of water and ions and become extremely dehydrated and fatigued because the loop of Henle is where the most reabsorption happens
    d. the person will be rushed to the hospital because he/she will be unable to reabsorb any amount of water or ions without the loop of Henle

11. If a person is given a diuretic: a substance that makes you pee more, how would you expect the collecting duct to change?
    a. it will contract and become smaller/narrower
    b. it will dilate and become bigger
    c. it will become less porous
    d. it will become more porous

12. If the ascending limb could not reabsorb ions, then the medulla next to the ascending limb will be _____ salty and the filtrate going into the distal convoluted would be ____ concentrated.
    a. Less; less
    b. Less; more
    c. More; less
    d. More; more

13. Why is kidney failure so dangerous? How does it interfere with our body's health?

14. Diabetes is one of the leading causes of kidney failure due to excess glucose in the blood. Although the glomerulus does not usually filter proteins from the blood, some proteins occasionally enter the filtrate when they bind with glucose molecules that carry them through. How would the presence of excess glucose in the blood affect the kidney's ability to filter blood and maintain homeostasis?

15. You are extremely dehydrated on a sunny day. You are given the choice of drinking either water, a sports drink, or a caffeinated drink of your choice. Recall that caffeine is a diuretic, and sports drinks contain electrolytes (which are ions). Which drink would keep you hydrated for a longer amount of time and why?

Appendix E

Posttest for Experiments 4 and 5

The relevant segment for each question is indicated in italics at the end of each question.

[Retention questions]

1. How does the kidney help maintain homeostasis and aid in osmoregulation? *(Segment 1)*
    a. The kidney regulates the amount of water in the blood, but not the amount of dissolved substances in it.
    b. The kidney regulates the amount of dissolved substances in the blood, but not the amount of water in it.
    c. The kidney regulates the amount of both water and dissolved substances in the blood.
    d. The kidney does not regulate the amount of water or ions and molecules in the blood, but simply passes it through a filtration system to remove toxins.

2. What is the difference between homeostasis and osmoregulation? *(Segment 1)*
    a. osmoregulation is one way the body maintains homeostasis
    b. homeostasis is one way the body maintains osmoregulation
    c. homeostasis and osmoregulation work together to help the kidney function
    d. the kidney is important for osmoregulation but not for homeostasis

3. Urea and uric acid are both substances that the body creates to dispose of: *(Segment 2)*
    a. Ammonia
    b. Glucose
    c. Filtrate
    d. Hemoglobin

4. What is NOT a difference between urea and uric acid? *(Segment 2)*
    a. mammals convert ammonia into urea, while birds, lizards and insects convert ammonia into uric acid
    b. urea needs to be dissolved in water and uric acid does not
    c. urea is less toxic in the blood than uric acid
    d. urea is basic and uric acid is acidic

5. What is NOT a function of the kidney? *(Segment 3)*
    a. to maintain the body's stable internal environment
    b. to regulate levels of water & substances in the body
    c. to metabolize byproducts of food
    d. to control blood pressure

6. How do the nephrons relate to the kidney? *(Segment 3)*
    a. the kidney contains millions of nephrons
    b. each kidney has only one nephron
    c. the nephron is about the same size as the kidney
    d. recycled nephrons make up the material of the kidney

7. What is the main function of the glomerulus and Bowman's capsule? *(Segment 4)*
    a. To squeeze some fluid, urea, and ions out of the blood and into the nephron
    b. To squeeze some proteins and red blood cells out of the blood and into the nephron
    c. To squeeze some blood containing urea and ions into the nephron
    d. To squeeze some blood containing proteins and red blood cells into the nephron

8. How are the glomerulus and Bowman's capsule related? *(Segment 4)*
    a. blood passes through both of them (glomerulus filters the blood and the Bowman's capsule holds the filtered blood)
    b. the glomerulus holds filtrate and the Bowman's capsule holds filtered blood
    c. the glomerulus sends the filtered blood back into the body and the Bowman's capsule takes in unfiltered blood
    d. the glomerulus filters the blood and the Bowman's capsule holds filtrate

9. Which substance is NOT reabsorbed in the proximal convoluted tubule? *(Segment 5)*
    a. glucose
    b. amino acids
    c. ions
    d. proteins

10. Which of the following is NOT true? *(Segment 6)*
    a. the renal medulla is very salty compared to renal cortex
    b. the glomerulus is located in the renal medulla
    c. both convoluted tubules are located in the renal cortex
    d. the loop of henle is located in the renal medulla

11. Which is NOT a function of the loop of henle? *(Segment 7)*
    a. to pump urea into the filtrate
    b. to create a concentration gradient to pull more water out of the filtrate
    c. to pump ions out of the filtrate
    d. to reabsorb water from the filtrate

12. What substances are reabsorbed in the descending limb? *(Segment 8)*
    a. sodium, potassium, chlorine
    b. urea and ammonia
    c. water only
    d. amino acids, glucose, and calcium

13. What substances are reabsorbed in the ascending limb? *(Segment 9)*
    a. sodium, potassium, chlorine
    b. urea and ammonia
    c. water only
    d. amino acids, glucose, and calcium

14. Which one of these statements about the distal convoluted tubule is NOT true? *(Segment 10)*
    a. it regulates potassium, calcium, and sodium
    b. it is the convoluted tubule furthest away from the glomerulus
    c. it uses a concentration gradient to reabsorb ions
    d. it uses pumps and hormones to regulate the absorption process

15. What is the function of the collecting duct? *(Segment 11)*
    a. to absorb urea and make urine
    b. to regulate the antidiuretic hormone
    c. to collect excess glucose and water to make urine
    d. to reabsorb more water before the urine goes to the bladder

16. How do the ureters, urethra, and bladder work together to excrete urine? *(Segment 12)*
    a. the urethra reabsorb water into the body and the ureters do not
    b. the urethra carries urine from the kidney into the bladder, and the ureter carries urine from the bladder out of the body
    c. the ureters reabsorb water into the body and the urethra does not
    d. the ureters carry urine from the kidney into the bladder, and the urethra carries urine from the bladder out of the body

[Transfer questions]

17. What would happen if we make the collecting duct's membrane more porous? *(Segment 11)*
    a. ions from the medulla will leak back into the collecting duct and the urine will be more concentrated
    b. more water from the filtrate will be reabsorbed into the medulla and the urine will be more concentrated
    c. both ions and water from the medulla will leak back into the collecting duct and the urine's color will not change but there will be more urine
    d. both ions and water will be reabsorbed into the medulla and the urine's color will not change but there will be less urine

18. What would happen to the amount of water and ions reabsorbed by the nephron if someone did not have a loop of Henle? *(Segment 7)*
    a. the person will not feel a difference because the amount of water and ions reabsorbed in the loop of Henle is nearly negligible
    b. the person will not feel a difference because the proximal convoluted tubule, distal convoluted tubule and collecting duct will provide enough reabsorption of water and ions
    c. the person will lose a lot of water and ions and become extremely dehydrated and fatigued because the loop of Henle is where the most reabsorption happens
    d. the person will be rushed to the hospital because he/she will be unable to reabsorb any amount of water or ions without the loop of Henle

19. If a person is given a diuretic: a substance that makes you pee more, how would you expect the collecting duct to change? *(Segment 11)*
    a. it will contract and become smaller/narrower
    b. it will dilate and become bigger
    c. it will become less porous
    d. it will become more porous

20. If the ascending limb could not reabsorb ions, then the medulla next to the ascending limb will be ____ salty and the filtrate going into the distal convoluted would be ____ concentrated. *(Segment 7)*
    a. Less; less
    b. Less; more
    c. More; less
    d. More; more

Appendix F

Posttest for Experiments 6 and 7

[Retention Questions]

1. How does the kidney filter blood? You can use the word bank below to help you describe the process:
    - nephron
    - concentration gradient
    - osmosis
    - water
    - glucose
    - ions
    - urea
    - filtrate
    - glomerulus
    - bowman's capsule
    - loop of Henle
    - collecting duct
    - proximal & distal convoluted tubes
    - descending & ascending limbs

2. What is the difference between homeostasis and osmoregulation?
    a. osmoregulation is one way the body maintains homeostasis
    b. homeostasis is one way the body maintains osmoregulation
    c. homeostasis and osmoregulation work together to help the kidney function
    d. the kidney is important for osmoregulation but not for homeostasis

3. Urea and uric acid are both substances that the body creates to dispose of:
    a. Ammonia
    b. Glucose
    c. Filtrate
    d. Hemoglobin

4. About ____ of the filtrate is reabsorbed back into the body and about ____ is excreted as urine:
    a. 99%, 1%
    b. 90%, 10%
    c. 75%, 25%
    d. 50%, 50%

5. The glomerulus filters the blood from the ______ and the resulting filtrate collects in the _______ before it flows through the rest of the nephron.
    a. renal vein, proximal convoluted tubule
    b. renal artery, Bowman's capsule
    c. renal vein, Bowman's capsule
    d. renal artery, proximal convoluted tubule

6. Which substance is NOT reabsorbed in the proximal convoluted tubule?
    a. glucose
    b. amino acids
    c. ions
    d. proteins

7. Which structure is located in the renal medulla?
    a. Glomerulus
    b. Proximal convoluted tubule
    c. Loop of Henle
    d. Distal convoluted tubule

8. Which is NOT a function of the loop of Henle?
    a. to pump urea into the filtrate
    b. to create a concentration gradient to pull more water out of the filtrate
    c. to pump salts out of the filtrate
    d. to reabsorb water from the filtrate

9. What substances are reabsorbed in the descending limb of the loop of Henle?
    a. sodium, potassium, chlorine
    b. urea and ammonia
    c. water only
    d. amino acids, glucose, and calcium

10. What substances are reabsorbed in the ascending limb of the loop of Henle?
    a. sodium, potassium, chlorine
    b. urea and ammonia
    c. water only
    d. amino acids, glucose, and calcium

11. Which of the following statements about the distal convoluted tubule is FALSE?
    a. it regulates potassium, calcium, and sodium
    b. it is the convoluted tubule furthest away from the glomerulus
    c. it uses a concentration gradient to reabsorb ions
    d. it uses pumps and hormones to regulate the absorption process

12. What is a primary function of the collecting duct?
    a. to facilitate the reabsorption of urea
    b. to regulate the antidiuretic hormone
    c. to collect excess glucose
    d. to facilitate the final reabsorption of water

13. Which structure carries urine from the kidneys to the bladder?
    a. Distal convoluted tubule
    b. Collecting duct
    c. Ureter
    d. Urethra

[Transfer questions]

14. Imagine you are extremely dehydrated on a sunny day. You are given the choice of drinking either water, a sports drink, or a caffeinated drink of your choice. Recall that caffeine is a diuretic, and sports drinks contain electrolytes (which are ions). Which drink would rehydrate you most quickly and why?

15. What symptom would you NOT expect if your body became incapable of osmoregulation?
    a. high blood pressure
    b. swelling of hands and feet
    c. changes in urination frequency
    d. fever

16. What would happen to the amount of urea in your urine if you fasted for a week?
    a. It would increase
    b. It would decrease
    c. It would stay the same
    d. It's impossible to tell

17. Which is the best analogy for how the kidney filters the blood?
    a. Looking carefully through a drawer to pick out the things you want to throw out
    b. Looking carefully through a drawer to pick out the things you want to keep
    c. Dumping out a drawer and putting almost everything back into it except the few things you want to throw out
    d. Dumping out a drawer and putting back only a few things before throwing everything else out

18. If the pressure in the glomerulus decreased _____
    a. less fluid would be filtered out of blood
    b. no filtrate would be produced
    c. blood would flow into the Bowman's capsule at a slow rate
    d. blood would flow out of Bowman's capsule at a slow rate

19. If the proximal convoluted tubule and the distal convoluted tubule were to switch functions, how would the concentration of organic solutes in the loop of Henle change?
    a. The concentration would be lower
    b. The concentration would be higher
    c. The concentration would be the same
    d. It's impossible to tell

20. Why does most water reabsorption happen in the renal medulla?
    a. Because the concentration of salts is highest in the medulla
    b. Because the concentration of proteins is highest in the medulla
    c. Because that is the only place where the nephron is permeable to water
    d. Because water is actively pumped out of the nephron in the medulla

21. How is the length of the loop of Henle related to urine production?
    a. Longer loops lead to more water reabsorption
    b. Longer loops lead to less water reabsorption
    c. The length of the loop does not affect water reabsorption
    d. It depends on what kind of animal it is

22. If the renal medulla were less hypertonic (salty) what would happen to the water reabsorbed in the descending limb?
    a. less water would be reabsorbed because the concentration gradient would be lower
    b. more water would be reabsorbed because the concentration gradient would be higher
    c. the same amount of water would be reabsorbed because the concentration gradient would be the same
    d. it's impossible to tell

23. If the ascending limb did not have channels to facilitate active transport of ions from the filtrate to the medulla, the amount of urine produced would _____ and the color of the urine would be _____.
    a. increase, lighter than normal
    b. increase, darker than normal
    c. stay the same, lighter than normal
    d. stay the same, darker than normal

24. A change in hormones is most likely to affect reabsorption in which part of the nephron?
    a. the proximal convoluted tubule
    b. the descending limb of the loop of Henle
    c. the ascending limb in the loop of Henle
    d. the distal convoluted tubule

25. What would happen if the collecting duct's membrane became more porous?
    a. ions from the medulla would leak back into the collecting duct
    b. more water from the filtrate would be reabsorbed into the medulla
    c. both ions and water from the medulla would leak back into the collecting duct
    d. both ions and water from the filtrate would be reabsorbed into the medulla

26. A blockage in the ureter would prevent urine from traveling from the ___ to the ____
    a. kidney; bladder
    b. bladder; kidney
    c. urethra; bladder
    d. bladder; urethra

Appendix G

Crash Course memory outline

1. Clive Wearing example
    a. He was an accomplished London musician until he contracted herpesviral encephalitis.
    b. The disease ravaged his central nervous system, resulting in one of the most profound cases of chronic amnesia ever recorded.
        i. While Mr. Wearing was perfectly able to play the piano, speak English, and dress himself, he was unable to remember much of his past or form any new memories.
        ii. His wife was the only person he was able to recognize, though he could not recall the last time he saw her.
    c. His case shows that some kinds of memory can be damaged while other kinds still work just fine

2. Memory is learning that has persisted over time
    a. Information that has been stored and can be recalled at a later date
    b. 3 ways to access memories
        i. Recall: reach back in your memory and bring up a specific piece of information
            1. i.e. fill-in-the blank style questions
        ii. Recognition: identify old information when it is presented to you
            1. i.e. multiple choice style questions
        iii. Relearning: refreshing or reinforcing old information
            1. i.e. studying for a final exam and relearning items that you had previously forgotten

3. Atkinson-Shiffrin model
    a. The information processing model is a *conceptual* model of memory
        i. Brains are similar to computers
        ii. Input information from the environment and output decisions/behaviors
        iii. Doesn't describe *where* things happen in the brain
    b. Step 1: Sensory memory/sensory register
        i. Information from the environment is encoded in an immediate, but fleeting memory
        ii. 2 most frequently studied forms of sensory memory:
            1. Iconic (what you see)
                a. Incredibly vivid/detailed
                b. Typically lasts less than half a second
            2. Echoic (what you hear)
                a. Typically lasts about 3 to 4 seconds
    c. Step 2: Short-term memory/working memory
        i. At any given moment, we are exposed to more information than we can possibly process at once
            1. We choose what to pay attention to, and this information is passed along to working memory

        ii. Essentially, working memory includes whatever you are thinking about at the current moment
        iii. Capacity is 7±2 items/pieces of information
            1. Interesting fact: this is the reason why phone numbers started out 7 digits in length
        iv. You can "hold" items in your short term memory through rehearsal, a specific type of encoding
            1. For instance, repeating a phone number over and over to yourself
        v. Short-term memory has roughly a 30 second limit, after which the memory either decays or is passed along to long-term memory
    d. Step 3: Long term memory
        i. The final stage of the information processing model
        ii. Once information reaches long term memory, it is similar to hitting the "Save" button on your computer
        iii. You can think of long term memory as your brain's "durable and ridiculously spacious storage unit"
        iv. Capacity is unlimited (different from *processing capacity*)
        v. Contains all your knowledge, skills, and experiences

4. Working memory
    a. A modern update on Atkinson and Shiffrin's model of short term memory
    b. Components of working memory:
        i. Visuospatial sketchpad
            1. Processes visual and spatial information, like pictures and maps
        ii. Phonological loop
            1. Processes verbal information, such as words and numbers
            2. Used when you repeat a phone number over and over to yourself in an effort to remember it
        iii. Central executive
            1. Considered the " traffic cop" of working memory, the central executive is used to coordinate information requiring both the phonological loop and the visuospatial sketchpad
            2. For instance, a map with both street names (verbal information) and landmarks (visual information)
            3. The central executive creates an integrated representation of the information, which is passed along to the episodic buffer
        iv. Episodic buffer
            1. Stores information from working memory and acts as a connector to long-term memory
    c. A more comprehensive definition of working memory includes all the ways in which we take short-term information and store it in long-term memory. This includes:
        i. Implicit processes: automatic processes that do not require any active concentration
            1. i.e. classically conditioned responses, such as becoming nervous and sweating profusely upon visiting the dentist
        ii. Explicit processes: when we store information consciously and actively
            1. i.e. Studying for an exam

5. Different types of long-term memory
    a. Implicit (non-declarative) memories are memories you may not be able to explicitly articulate
        i. Procedural memory: how we do things, such as riding a bike
            1. May be effortful to lean at first, but eventually becomes automatic
        ii. Priming: when previous experience influences your interpretation of an event
    b. Explicit (declarative) memories are facts or events that you can clearly and explicitly describe
        i. Semantic: "Having to do with words"
        ii. Episodic: memory for specific episodes of your life
            1. Thus, while Clive Wearing's procedural memory appeared to be completely intact, his episodic memories were deeply impaired

6. Tricks for improving your memory
    a. Mnemonics can help to organize large amount of information into more familiar, manageable bits
        i. Acronyms (i.e. ROY G. BIV can help when remembering the colors of the rainbow)
    b. Levels of processing
        i. Shallow processing allows you to encode information on basic auditory and visual levels, such as the sound or font of a word
        ii. Deep processing allows you to encode information semantically, or based on the actual meaning of the word

Appendix H

Khan Academy memory outline

1. Information Processing Model
    a. Memory is learning that has persisted over time
        i. Information that has been stored and can be recalled at a later date
    b. The Atkinson-Shiffrin information processing model is a *conceptual* model of memory that uses a common analogy:
        i. Brains are similar to computers
        ii. Input information from the environment and output decisions/behaviors
        iii. Doesn't describe *where* things happen in the brain

2. Step 1: Sensory memory/sensory register
    a. A temporary register of all the information being taken in by the 5 senses
        i. Information from the environment is encoded in an immediate, but fleeting memory
    b. 2 most frequently studied forms of sensory memory:
        i. Iconic (what you see)
            1. Incredibly vivid/detailed
            2. Typically lasts less than half a second
        ii. Echoic (what you hear)
            1. Typically lasts about 3 to 4 seconds
    c. At any given moment, we are exposed to more information than we can possibly process at once
        i. We choose what to pay attention to, and this information is passed along to working memory

3. Step 2: Working memory/short-term memory
    a. Essentially, working memory includes whatever you are thinking about at the current moment
        i. Working memory is a modern update to Atkinson and Shiffrin's idea of "short-term" memory
    b. Capacity is 7±2 items/pieces of information
        i. Interesting fact: this is the reason why phone numbers started out 7 digits in length
        ii. Short-term memory has roughly a 30 second limit, after which the memory either decays or is passed along to long-term memory
        iii. Mnemonics can help to organize large amount of information into more familiar, manageable bits
            1. Acronyms (i.e. ROY G. BIV can help when remembering the colors of the rainbow)
    c. Components of working memory:
        i. Visuospatial sketchpad
            1. Processes visual and spatial information, like pictures and maps
        ii. Phonological loop
            1. Processes verbal information, such as words and numbers
            2. You can "hold" items in your phonological loop through rehearsal, a specific type of encoding

        a. i.e. when you repeat a phone number over and over to yourself in an effort to remember it
    iii. Central executive
        1. Considered the " traffic cop" of working memory, the central executive is used to coordinate information requiring both the phonological loop and the visuospatial sketchpad
        2. For instance, a map with both street names (verbal information) and landmarks (visual information)
        3. The central executive creates an integrated representation of the information, which is passed along to the episodic buffer
    iv. Episodic buffer
        1. Stores information from working memory and acts as a connector to long-term memory
  d. A more comprehensive definition of working memory includes all the ways in which we take short-term information and store it in long-term memory. This includes:
    i. Implicit processes: automatic processes that do not require any active concentration
        1. i.e. classically conditioned responses, such as becoming nervous and sweating profusely upon visiting the dentist
    ii. Explicit processes: when we store information consciously and actively
        1. i.e. Studying for an exam
  e. Many modern researchers have noted differences in how we process new information, also known as levels of processing
    i. Shallow processing allows you to encode information on basic auditory and visual levels, such as the sound or font of a word
    ii. Deep processing allows you to encode information semantically, or based on the actual meaning of the word

4. Step 3: Long-term memory
  a. The final stage of the information processing model
    i. You can think of long term memory as your brain's "durable and ridiculously spacious storage unit"
    ii. Contains all your knowledge, skills, and experiences
  b. Once information reaches long term memory, it is similar to hitting the "Save" button on your computer
  c. 3 ways to access information stored in long term memory
    i. Recall: reach back in your memory and bring up a specific piece of information
        1. i.e. fill-in-the blank style questions
    ii. Recognition: identify old information when it is presented to you
        1. i.e. multiple choice style questions
    iii. Relearning: refreshing or reinforcing old information
        1. i.e. studying for a final exam and relearning items that you had previously forgotten
  d. 2 main categories of memories:
    i. Explicit (declarative) memories are facts or events that you can clearly and explicitly describe
        1. Semantic memory: "Having to do with words"
        2. Episodic memory: memory for specific episodes of your life

            a. Thus, while Clive Wearing's procedural memory appeared to be completely intact, his episodic memories were deeply impaired
    ii. Implicit (non-declarative) memories are memories you may not be able to explicitly articulate
        1. Procedural memory: how we do things, such as riding a bike
            a. May be effortful to lean at first, but eventually becomes automatic
        2. Priming: when previous experience influences your interpretation of an event
e. Capacity is unlimited (different from *processing capacity*)

5. Clive Wearing example
    a. He was an accomplished London musician until he contracted herpesviral encephalitis.
    b. The disease ravaged his central nervous system, resulting in one of the most profound cases of chronic amnesia ever recorded.
        i. While Mr. Wearing was perfectly able to play the piano, speak English, and dress himself, he was unable to remember much of his past or form any new memories.
        ii. His wife was the only person he was able to recognize, though he could not recall the last time he saw her.
    c. His case shows that some kinds of memory can be damaged while other kinds still work just fine

Appendix I

Posttest used in Experiment 8

Retention Open Response:

1. How does information get into long-term memory? Describe the process in as much detail as you can:

2. What different kinds of memory are there? What makes them different from each other?

Retention Multiple Choice:

3. Information from the outside world is initially processed in your _____.
    a. Short term memory
    b. Phonological loop
    c. Working memory
    d. Sensory memory

4. Your _____ can hold around 7 + or – 2 bits of information at a time and lasts roughly _____ without rehearsal.
    a. Short term memory, 60 seconds
    b. Short term memory, 30 seconds
    c. Sensory register, 60 seconds
    d. Sensory register, 30 seconds

5. _____ memories are those that can be clearly articulated, while _____ memories are those that cannot.
    a. Explicit, implicit
    b. Implicit, explicit
    c. Declarative, semantic
    d. Semantic, episodic

6. Implicit memories for how to carry out actions and do things are _____ memories.
    a. Episodic
    b. Semantic
    c. Procedural
    d. Iconic

7. Explicit memories for events in your own life are _____ memories.
    a. Episodic
    b. Semantic
    c. Procedural
    d. Echoic

8. Semantic, episodic, and procedural memory are all types of _____.
    a. Sensory memory
    b. Short term memory
    c. Working memory
    d. Long term memory

Transfer Open Response:

9. How is memory in the brain similar to memory in a computer? How is it different?

10. What could you do to improve your long-term memory?

11. You are driving to a new location and you've memorized the directions. Before you left, you studied a map and you also reviewed the turn-by-turn directions. How do you represent that information in your working memory?

12. Do you think all memories are stored in one spot in the brain in many spots? Why?

Transfer Multiple Choice:

13. Your mom is lecturing you about a chore you forgot to do, but you are not really listening. Suddenly she says, "Are you listening?? What did I just say?" and you are able to repeat back the last thing she said. What kind of memory did you use to do this?
    a. Sensory
    b. Short-term
    c. Working
    d. Long-term

14. You are taking an exam and you come to a question that you <u>feel</u> you know – you can picture the class when it was covered and where in your notebook it is – but you can't come up with the answer. What kind of memory are you experiencing?
    a. Procedural
    b. Priming
    c. Semantic
    d. Episodic

15. Sarah received her driver's license in California when she was 16. After 2 years of driving on a daily basis, Sarah moved to New York City where she relies solely on public transportation. When Sarah returns home for a visit and borrows her parents' car, she discovers that, despite not having driven for over 4 years, she still remembers exactly how it is done. This is an example of a(n) _____ memory.
    a. Semantic
    b. Episodic
    c. Procedural
    d. Explicit

16. For Melissa's 6th birthday, her parents threw her a lavish surprise party complete with a petting zoo and puppy pen. This birthday party was so memorable that even at 24, Melissa can still recall the day's events with vivid detail. This is an example of a(n) _____ memory.
    a. Procedural
    b. Semantic
    c. Episodic
    d. Implicit

17. Patient HM suffered from chronic epilepsy. In order to ease the severity of his symptoms, a team of surgeons removed the portion of his brain responsible for generating the seizures. After the procedure, HM suffered from profound memory impairment. For instance, while he was able to remember information for short periods of time (around 30 seconds), HM had lost the capacity to form any new, lasting memories. Psychologists would say that while his _____ is intact, his _____ is severely impaired.
    a. short-term memory; long-term memory
    b. short-term memory; working memory
    c. episodic memory; working memory
    d. working memory, sensory memory

18. Which of the following question types is most like a “recognition” memory test?
    a. Fill in the blank
    b. True-false
    c. Short answer
    d. Multiple choice

References

Anderson, J. R., & Reder, L. M. (1979). An elaborative processing explanation of depth of processing. *L.; S. Cermak and FIM Craik, Eds., Levels of Processing in Human Memory (Erlbam, 1979)*, 385-404.

Ausubel, D. P. (1960). The use of advance organizers in the learning and retention of meaningful verbal material. *Journal of educational psychology*,*51*(5), 267.

Ausubel, D. P. (1978). In defense of advance organizers: A reply to the critics.*Review of Educational research*, 251-257.

Bjork, E. L., & Bjork, R. A. (2011). Making things hard on yourself, but in a good way: Creating desirable difficulties to enhance learning. *Psychology and the real world: Essays illustrating fundamental contributions to society*, 56-64.

Bransford, J. D. & Schwartz, D. L. (1999). Rethinking transfer: A simple proposal with multiple implications. *Review of Research in Education, 24,* 61-100.

Britton, B. K., & Gülgöz, S. (1991). Using Kintsch's computational model to improve instructional text: Effects of repairing inference calls on recall and cognitive structures. *Journal of Educational Psychology*, *83*(3), 329.

Britton, B. K., Van Dusen, L., Gulgöz, S., & Glynn, S. M. (1989). Instructional texts rewritten by five expert teams: Revisions and retention improvements. *Journal of Educational Psychology*, *81*(2), 226.

Chi, M. T. (2009). Active-constructive-interactive: A conceptual framework for differentiating learning activities. *Topics in Cognitive Science*, *1*(1), 73-105.

Chi, M. T., Bassok, M., Lewis, M. W., Reimann, P., & Glaser, R. (1989). Self-explanations: How students study and use examples in learning to solve problems. *Cognitive science*, *13*(2), 145-182.

Chi, M. T., Leeuw, N., Chiu, M. H., & LaVancher, C. (1994). Eliciting self-explanations improves understanding. *Cognitive science*, *18*(3), 439-477.

Craik, F. I., & Lockhart, R. S. (1972). Levels of processing: A framework for memory research. *Journal of verbal learning and verbal behavior*, *11*(6), 671-684.

Craik, F. I., & Tulving, E. (1975). Depth of processing and the retention of words in episodic memory. *Journal of experimental Psychology: general*, *104*(3), 268.

Glaser, R. (1991). The maturing of the relationship between the science of learning and cognition and educational practice. *Learning and instruction*, *1*(2), 129-144.

Glaser, R., & Bassok, M. (1989). *Learning theory and the study of instruction* (No. UPITT/LRDC/ONR/KBC-12). PITTSBURGH UNIV PA LEARNING RESEARCH AND DEVELOPMENT CENTER.

Glenberg, A. M., & Epstein, W. (1985). Calibration of comprehension. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *11*(4), 702.

Glenberg, A. M., Sanocki, T., Epstein, W., & Morris, C. (1987). Enhancing calibration of comprehension. *Journal of Experimental Psychology: General*,*116*(2), 119.

Harp, S. F., & Mayer, R. E. (1997). The role of interest in learning from scientific text and illustrations: On the distinction between emotional interest and cognitive interest. *Journal of Educational Psychology, 89,* 92–102.

Harp, S. F., & Mayer, R. E. (1998). How seductive details do their damage: A theory of cognitive interest in science learning. *Journal of Educational Psychology, 90,* 414–434.

Kapur, M. (2008). Productive failure. *Cognition and Instruction*, *26*(3), 379-424.

Kapur, M., & Bielaczyc, K. (2012). Designing for productive failure. *Journal of the Learning Sciences*, *21*(1), 45-83.

Kintsch, W. (1994). Text comprehension, memory, and learning. *American Psychologist*, *49*(4), 294.

Kintsch, W. (1988). The role of knowledge in discourse comprehension: a construction-integration model. *Psychological review*, *95*(2), 163.

Kintsch, W., & Van Dijk, T. A. (1978). Toward a model of text comprehension and production. *Psychological review*, *85*(5), 363.

Mannes, S. M., & Kintsch, W. (1987). Knowledge organization and text organization. *Cognition and instruction*, *4*(2), 91-115.

Mautone, P. D., & Mayer, R. E. (2001). Signaling as a cognitive guide in multimedia learning. *Journal of Educational Psychology, 93,* 377–389.

Mayer, R. E. (1979). Can advance organizers influence meaningful learning? *Review of educational research*, 371-383.

Mayer, R. E. (2005). Cognitive theory of multimedia learning. *The Cambridge handbook of multimedia learning*, 31-48.

Mayer, R. E. (2008). Applying the science of learning: evidence-based principles for the design of multimedia instruction. *American Psychologist*, *63*(8), 760.

Mayer, R. E., & Chandler, P. (2001). When learning is just a click away: Does simple user interaction foster deeper understanding of multimedia messages? *Journal of Educational Psychology, 93,* 390–397.

Mayer, R. E., Dow, G., & Mayer, S. (2003). Multimedia learning in an interactive self-explaining environment: What works in the design of agent-based microworlds? *Journal of Educational Psychology, 95,* 806–813.

Mayer, R. E., & Jackson, J. (2005). The case for coherence in scientific explanations: Quantitative details can hurt qualitative understanding. *Journal of Experimental Psychology: Applied, 11,* 13–18.

Mayer, R. E., Mathias, A., & Wetzell, K. (2002). Fostering understanding of multimedia messages through pretraining: Evidence for a two-stage theory of mental model construction. *Journal of Experimental Psychology: Applied, 8,* 147– 154.

Mayer, R. E., Mautone, P., & Prothero, W. (2002). Pictorial aids for learning by doing in a multimedia geology simulation game. *Journal of Educational Psychology, 94,* 171–185.

McNamara, D. S. (2001). Reading both high-coherence and low-coherence texts: Effects of text sequence and prior knowledge. *Canadian Journal of Experimental Psychology/Revue canadienne de psychologie expÃ© rimentale*,*55*(1), 51.

McNamara, D. S., & Kintsch, W. (1996). Learning from texts: Effects of prior knowledge and text coherence. *Discourse processes*, *22*(3), 247-288.

McNamara, D. S., Kintsch, E., Songer, N. B., & Kintsch, W. (1996). Are good texts always better? Interactions of text coherence, background knowledge, and levels of understanding in learning from text. *Cognition and instruction*, *14*(1), 1-43.

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

Rickards, J. P. (1976). Interaction of position and conceptual level of adjunct questions on immediate and delayed retention of text. *Journal of Educational Psychology*, *68*(2), 210.

Rickards, J. P., & Di Vesta, F. J. (1974). Type and frequency of questions in processing textual material. *Journal of Educational Psychology*, *66*(3), 354.

Rickards, J. P., & Hatcher, C. W. (1977). Interspersed meaningful learning questions as semantic cues for poor comprehenders. *Reading Research Quarterly*, 538-553.

Rothkopf, E. Z. (1966). Learning from written instructive materials: An exploration of the control of inspection behavior by test-like events. *American Educational Research Journal*, 241-249.

Rothkopf, E. Z., & Bisbicos, E. E. (1967). Selective facilitative effects of interspersed questions on learning from written materials. *Journal of Educational Psychology*, *58*(1), 56.

Roy, M., & Chi, M. T. (2005). The self-explanation principle in multimedia learning. *The Cambridge handbook of multimedia learning*, 271-286.

Schwartz, D. L., & Martin, T. (2004). Inventing to prepare for future learning: The hidden efficiency of encouraging original student production in statistics instruction. *Cognition and Instruction*, *22*(2), 129-184.

Sweller, J. (2005). The redundancy principle in multimedia learning. In R. E. Mayer (Ed.), *The Cambridge handbook of multimedia learning* (pp. 159-168). New York: Cambridge University Press.

Sweller, J., van Merrienboer, J. J. G., & Paas, F. G. W. C. (1998). Cognitive architecture and instructional design. *Educational Psychology Review, 10*, 251-296.

Walczyk, J. J., & Hall, V. C. (1989). Effects of examples and embedded questions on the accuracy of comprehension self-assessments. *Journal of Educational Psychology*, *81*(3), 435.