Perspectives from the CHDS Media Hub
Cognitive Theory of Multimedia Learning

Jacob B Waxman, Ed.M., A.L.M.
Sue J. Goldie, MD, MPH
Center for Health Decision Science, Harvard T.H. Chan School of Public Health

Cognitive Theory of Multimedia Learning

Applying cognitive load theory to the design of educational multimedia

Introduction

The cognitive theory of multimedia learning (CTML) draws together various theories of learning and cognition- particularly Paivio's dual encoding theory (Clark & Paivio 1991) and cognitive load theory (CLT) (Sweller et al. 1998) – and applies them to the process of learning through multimedia instruction. At its core, the theory asserts that people learn more effectively from words and graphics than from words alone. While numerous investigators have contributed to this work, we have found Mayer's (2009) empirically derived principles of multimedia design, to be the most relevant and useful for instructional designers and educators.

What is multimedia instruction?

It is not a video! Multimedia instruction (also referred to as "educational multimedia") refers to teaching and learning material that contain words (e.g., spoken or printed text) and images (e.g., illustrations, charts, photos, animation, video) (Mayer 2009). Multimedia instruction therefore has relevance well beyond a video-based lesson – it can inform our teaching over a broad range of visual and audio modalities, from what we sketch on a chalkboard while lecturing to how we construct a diagram we create for a handout in class, or the design of a power-point slide to use in class.

Where did the cognitive theory of multimedia learning come from?

The CTML, extensively described by Mayer (2009, 2011, 2014), is based on CLT in an environment of multimedia learning. CLT is a theory of learning that is based on what we know about how learners process new information and construct knowledge in long-term memory (Sweller 1998, 2020). See our monograph on "Cognitive Load Theory and Instructional Design" (Goldie and Waxman 2021) for a review of the theory, assumptions, and implications for instructional design.

What is cognitive load?

Cognitive load is a central concept critical to cognitive load theory and highly relevant to learning and instructional design (Sweller 1994, 1999). Cognitive load refers to the "cognitive processing demands" placed on a learner in the context of the limitations of working memory. There are three general types of cognitive load that have been described: (1) intrinsic cognitive load, (2) germane cognitive load, and (3) extraneous cognitive load.

1. <u>Intrinsic cognitive load</u>

Intrinsic cognitive load refers to the cognitive processing that is needed to comprehend new information. Intrinsic cognitive load will depend on the nature of the material, the learner's prior knowledge, the learner's affect, and contextual features of the learning environment (Clark 2006). Framed in terms of cognitive processing, Mayer describes the selecting of relevant information and organizing it in working memory as "***essential processing***" (Mayer 2009).

2. <u>Germane cognitive load</u>

Germane cognitive load refers to intrinsic cognitive load that is associated with more difficult learning tasks, more complex learning material, and occurs when learners engage in "deep cognitive processing" (Mayer 2008) and quite related to a learner's effort. Framed in terms of cognitive processing, Mayer describes the "making sense of new material by reorganizing it into a coherent structure and integrating it with relevant prior knowledge," as ***"generative processing"*** (Mayer 2009).

3. <u>Extraneous cognitive load</u>

Unlike intrinsic and germane cognitive load, which are both essential for learning, extraneous cognitive load refers to information processing that is *not needed* or irrelevant to comprehend the new material in a lesson plan. This might include (a) unrelated information, such as visual elements with no purpose, (b) extensive details that are not necessary for the particular learning task, (c) an instructional format that introduces distraction, such as animation for "animation" sake. Framed in terms of cognitive processing, Mayer refers to the draining of limited cognitive processing capacity without contributing to learning as "***extraneous processing***" (Mayer 2009).

Note that all three types of cognitive load need to be considered in the context of (a) learner expertise, (b) content complexity, and (c)

Cognitive Theory of Multimedia Learning (cont.)

instructional methods and pedagogical strategies.

How can we design educational multimedia that contributes to effective learning through attention to cognitive processing?

In a nutshell, the goal is to make the cognitive processes associated with learning as effective and efficient as possible, using a three-pronged strategy: (1) Reducing extraneous processing, (2) Managing essential processing (3) Fostering generative processing. Just as Mayer adapted the cognitive load theory to multimedia learning, he contextualized the goals above using principles that were established based on empirical data (Mayer 2004, 2008, 2009, 2012) We expand on these three strategies below.

1. Reducing Extraneous Processing

Learners have a limited WM capacity; therefore, we want to be efficient with how we utilize it! Intrinsic and germane cognitive loads have pedagogical value, while extraneous cognitive load has none- therefore we should aim to first reduce any extraneous load on the learner. Mayer (2009) identified five principles, based on empirical evidence, to reduce extraneous overload in the context of educational multimedia: (a) Coherence, (b) Redundancy, (c) Signaling, (d) Spatial contiguity, (e) Temporal contiguity.

The Coherence Principle

"People learn more deeply from a multimedia message when extraneous material is excluded rather than included" (Mayer 2009). Eliminating distracting and unnecessary material (e.g., humorous graphics or irrelevant animations) results in better learning. Simplify learner workflows and eliminate unnecessary content.

The Redundancy Principle

"People learn more deeply from graphics and narration than from graphics, narration, and on-screen text" (Mayer 2009). Learners may focus on the printed words rather than the relevant portions of the graphics and can expend limited processing capacity in WM by trying to reconcile the two verbal streams of information.

The Signaling Principle

"People learn more deeply from a multimedia message when cues are added that highlight the organization of essential material" (Mayer 2009). Figuring out what to pay attention to counts as cognitive load! Signaling what's important, through visual organization and cues (e.g., outline, headings, highlighting) can reduce the attentive and cognitive load on the learner and free up working memory space for intrinsic or germane cognitive processing.

The Spatial Contiguity Principle

"People learn more deeply from a multimedia message when corresponding printed words and graphics are presented near rather than far from each other on the page or screen" (Mayer 2009). Presenting information close to its media counterparts reduces the amount of time and cognitive effort the learner must expend to separate out and connect the relevant pieces of information, freeing up capacity for intrinsic or germane cognitive processing.

The Temporal Contiguity Principle

"People learn more deeply from a multimedia message when corresponding graphics and narration are presented simultaneously rather than successively" (Mayer 2009). Sensory and WM are limited in terms of both time and capacity! We can only hold information in these memories for mere seconds. When complimentary pieces of information are presented near each other in time, the learner can spend more of that time capacity working to organize the new information and assimilate knowledge.

The Image Principle

"People do not necessarily learn more deeply from a multimedia message when the speaker's image is visible on screen (Mayer 2009). Every instant of a multimedia message is information that must either be assimilated or rejected by WM- the teacher's face isn't always integral to learning the material and doesn't necessarily result in greater learning. This should be weighed with affective considerations.

1. Reduce Extraneous Processing

- **Coherence:** Delete extraneous material
- **Signaling:** Highlight essential material
- **Redundancy:** Don't add onscreen captions to narrated graphics
- **Spatial contiguity**: Place printed words near corresponding part of graphic
- **Temporal contiguity:** Present spoken words at same time as corresponding graphics
- **Image Principle:** *People don't necessarily learn better when the speaker's image is visible.*

2. Manage Essential Processing

- **Segmenting:** *Break lesson into learner-paced parts*
- **Pre-training:** *Present characteristics of key concepts before lesson*
- **Modality:** *Use spoken words rather than printed words*

3. Foster Generative Processing

- **Personalization:** Put words in conversational style rather than formal style
- **Voice:** Put words in human voice rather than machine voice
- **Embodiment:** Have onscreen agent use human-like gestures and movements

(Mayer 2001)

Cognitive Theory of Multimedia Learning (cont.)

2. Managing Essential Processing
Essential overload occurs the cognitive capacity of the learner is exceeded by the amount of essential processing a new piece of information requires. Mayer (2009) identified three principles, based on empirical evidence, to manage essential processing in the context of educational multimedia: (a) Segmenting, (*b*) Pre-training, and (c) Modality.

The Segmenting Principle
"People learn more deeply when a multimedia message is presented in learner-paced segments rather than as a continuous unit" (Mayer 2009).
Segmenting allows learners to fully process one "chunk" of the learning experience, before having to move onto the next one. If we chunk our content into manageable chunks (defined by the content complexity and learner level), learners can more efficiently manage essential cognitive processing in WM.

The Pre-training Principle
"People learn more deeply from a multimedia message when they have learned the names and characteristics of the main concepts" (Mayer 2009).
Providing students with definitions and main characteristics of key elements in a lesson, frees up cognitive capacity for more important essential cognitive tasks, such as understanding the relationships between pieces of information, integrating new information into existing schema, and creating new schema.

The Modality Principle
"People learn more deeply from a multimedia message when the words are spoken rather than printed" (Mayer 2009) Spreading complimentary information across the verbal and visual channels can be more efficient than using just one of them. Consider a diagram with adjacent text versus voice narration. With voice narration, the visual channel can be used to process the diagram, while the verbal channel can be used for the information that would have been in the adjacent text.

3. Fostering Generative Processing
Principles aimed at fostering generative processing "use social cues to prime the learner's motivation to exert effort to make sense of the material" (Mayer 2009). Mayer (2009) identified three principles, based on empirical evidence, to foster generative processing in the context of educational multimedia: (a) Personalization, (b) Voice, (c) Embodiment.

The Personalization Principle
"People learn better from multimedia lessons when words are in conversational style rather than formal style " (Mayer 2009)
An informal conversational style established an environment of "social proximity" and "social presence" between educator and learner. This both supports the necessary "essential" cognitive processing but also contributes to enhancing learner motivation to engage meaningfully and participate in active learning.

The Voice Principle
"People learn better when the narration in multimedia lessons is spoken in a friendly human voice rather than a machine voice" (Mayer 2009).
Similar to the personalization principle, a human voice contributes to "social proximity" and "social presence" between educator and learner.

The Embodiment Principle
"We learn more effectively when the instructors on screen in a multimedia lesson demonstrate movement, eye contact, facial expression and convey genuine human connectedness through physical gestures" (Mayer 2009).
This eye contact and physical expression of "connectedness" contributes to "social presence."

References

Clark, J. M. & Paivio, A. (1991). Dual coding theory and education. Educational Psychology Review, 3(3), 149-170.

Mayer RE (2001). Multimedia learning. New York: Cambridge University Press.

Mayer RE (2008). Applying the science of learning: Evidence-based principles for the design of multimedia instruction. Cognition and Instruction 19, 177-213.

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

Miller, G. (1956). The magical number seven, plus or minus two: Some limits on our capacity for processing information. The psychological review, 63, 81-97.

Paivio, A. (1986). Mental Representations. New York: Oxford University Press.

Sweller J (1988). Cognitive load during problem solving: Effects on learning. Cognitive Science 12, 257-285.

Sweller J, van Merrienboer JJG, Paas F. Cognitive architecture and instructional design. Ed Psychol Rev. 1998;10(3):251–296.

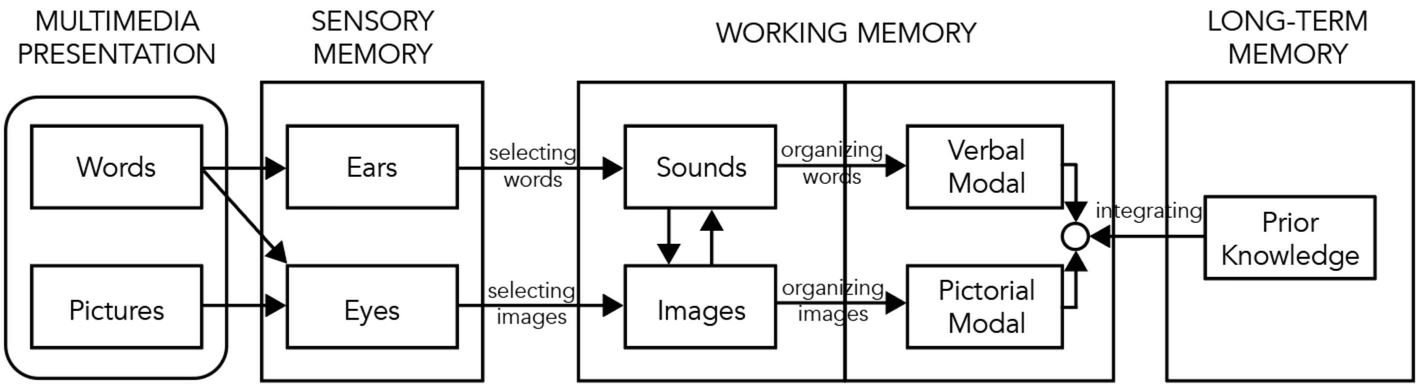
Sweller J (1989). Cognitive technology: Some procedures for facilitating learning and problem-solving in mathematics and science. Journal of Educational Psychology 81, 457-466.

Sweller J (1994). Cognitive load theory, learning difficulty, and instructional design. Learning and Instruction 4, 295-312.

Thomas M, et al. Annual Research Review: Educational neuroscience: progress and prospects. Journal of Child Psychology and Psychiatry 60:4 (2019), pp 477–492.

Cognitive Theory of Multimedia Learning (cont.)

The figure below "represents memory stores as rectangles: ***sensory memory***, which temporarily holds incoming images and sounds; ***working memory***, which allows for mentally manipulating a small amount of the incoming visual and verbal material; and ***long-term memory***, which is the learner's permanent storehouse of knowledge" (Mayer 2009).



"Cognitive processing includes (1) *selecting*, which transfers some of the incoming images and sounds to working memory for additional processing; (2) *organizing*, which organizes the images into a pictorial model and the words into a verbal model in working memory; and (3) *integrating*, which connects the models with each other and with relevant knowledge activated from long-term memory" (Mayer 2009).

Center for Health Decision Science
CHDS
Harvard T.H. Chan School of Public Health