Abstract: Instructional multimedia has rapidly advanced in the past decade. The area that has seen the most growth is video instruction. In this literature review the Cognitive Theory of Multimedia Learning (CTML) is introduced and how it can be implemented to help learners effectively and efficiently learn through instructional multimedia. The theoretical framework of CTML is reviewed as well has how memory works in instructional multimedia. The framework for multimedia learning is rooted in three assumptions: dual channels, limited capacity, and active processing. To balance visual and verbal channels the twelve principles of multimedia design are used. Research and experiments have been conducted by Mayer as well as other researchers over the past three decades on how implementing different multimedia principles can have an effect on a learner's ability to develop meaningful learning and what principles are the most effective in designing instructional multimedia. The results found that learners perform better on problem solving transfer tests when watching a concise lesson compared to an extended lesson. Text should be close to images and given breathing room. Using symbols and highlights help call out key concepts to the learner. Results also showed extraneous details should not be added.

Keywords: Cognitive Theory of Multimedia Learning, instructional multimedia, meaningful learning, dual channels, limited capacity, active processing, and multimedia principles
students learn by using graphics (graphs, photos, maps, animations, and videos) and printed or spoken text (Mayer, 2008). Graphics, photos, maps, text, and spoken words can be combined into one piece or used separately.

This paper will synthesize research on what the Cognitive Theory of Multimedia Learning is, what is meaningful learning, how multimedia learning works, the multimedia principles, the different types of multimedia presentations that can be created, how effective they are, and their limitations.

The Science of Learning

The science of learning investigates how people learn (Mayer, n.d). Understanding how people learn can assist instructional and multimedia designers as they developed sound instructional multimedia pieces that will generate meaningful learning.

What is meaningful learning? Mayer and Moreno (2003) defines meaningful learning as a deep understanding of the material. Learners emerge themselves in meaningful learning when they are able to make a connection between information in the visual and the verbal processing channels of working memory (Tempelman-Kluit, 2006). The learner is able to identify key concepts, mentally organize the information, and integrate this information with prior knowledge (Mautone & Mayer, 2001). A challenge all multimedia designers face is how to introduce new concepts that are engaging without causing cognitive overload. Cognitive overload according to Mayer and Moreno (2003) is "when the learner’s intended cognitive processing exceeds the learner’s available cognitive capacity (p. 43).

What is multimedia learning and instruction? Mayer and Moreno (2003) define multimedia learning as learning from words and pictures and multimedia instruction as the vehicle that delivers the words and pictures for learning (Mayer and Moreno). The words can be printed on the screen or spoken as a narration. The pictures can be static or animation. Static pictures can be charts, graphs, diagrams, and illustrations. Animation pictures can be interactive animations and videos.

Cognitive Theory of Multimedia Learning

The core principle of the CTML is how does multimedia learning work? How does the learner make sense of the instructional material and construct meaningful connections and new knowledge (Sorden, 2012)?

Theoretical framework. Richard E. Mayer is cognitive scientist who developed this theory and has the spent almost three decades researching and updating this theory has instructional multimedia has evolved. Mayer built this multimedia model off the work of Palvio, Baddelev, and Sweller. His model and principle’s for multimedia design were created after conducting numerous experiences on students from the Psychology Subject Pool at the University of California, Santa Barbara (Reed, 2006). In his instructional multimedia experiments he tested different types of cognitive constraints on multimedia learning. Learners were split into different groups during the experiment. After the multimedia piece(s) were completed learners were asked
to answer on paper a series of questions that measured the retention of what was learned.

**Three assumptions.** The framework for multimedia learning and how the mind works is rooted in these three assumptions: dual channels, limited capacity, and active processing. These three assumptions as well as how they are activated in this model are explained below:

- **Dual channels:** according to Austin (2009) "the dual channel processing assumption is based on the seminal work by Paivio (p. 1340)". Learners have different channels in their brain for processing visual and verbal material separately (Mayer & Moreno, 2003). The learner will select relevant words for processing in verbal working memory and relevant images for processing in visual working memory (Toh, Munassar, & Yahaya. 2010).
- **Limited capacity:** there is a limit to the amount of information (verbal and visual) each channel can process.
- **Active processing:** in order for meaningful and deeper learning to occur it is dependent on the learner's cognitive processing to be able to select, organize, and integrated the information (verbal and visual) being presented with prior knowledge (Mayer, 2008).

**How memory works in instructional multimedia.** These three assumptions are connected to the cognitive theory of multimedia learning graphic (figure 1). This figure represents how memory works in instructional multimedia. There are two rows and five columns of boxes that have arrows that connect them. According to Mayer and Moreno (2003) "the two rows contain information-processing channels (auditory/verbal channel and then visual/pictorial channel) (p.44)". There are five columns in this model that represent the modes of knowledge presentation. Learners start by watching an instructional multimedia piece. The multimedia presentation contains words (text and/or auditory) and pictures. Words and pictures are the physical representations. Learners then use their ears and eyes to access the sensory representations. Learners than select the text/auditory elements and images to put into working memory. The learner determines what text/auditory and images will be stored in long term memory. In order for the learner to process and integrate written text and visualizations, the written text or visual that the learner looked at first needs to be held in working memory as the learner looks at the second source that was not attended to first (Schmidt-Weigand, Kohnert, & Glowalla, 2010). The learner will then integrate these selected elements (words and images) with relevant prior knowledge. The arrows in this model represent cognitive processing. The arrow from words to ears represents the processing of the spoken word by the ears; the arrow from words to eyes is for printed text. Pictures are processed by the eyes. Moving from the sensory memory to working memory the arrows selecting words and images indicate the learner is selecting specific words and images to pay attention to. Then these selected words and images are organized into a coherent verbal and pictorial presentation. The last arrow moves from working memory to long-term memory. This is where the learner merges the verbal and pictorial model with relevant prior knowledge. Working memory is limited in storage and is temporally, whereas long-term memory has no limitations (Schweppe & Rummer, 2014).
Cognitive load theory. Cognitive Load Theory (CLT) defines how the brain can only process selective incoming sensory data into working memory (Sorden, 2005). This is an important theory multimedia and instructional designers must follow when designing instructional multimedia. Because learners can only process a limited amount of information at a time the information presented to the learner should not contain unnecessary content. Examples of unnecessary content are entertaining animation designs that distract the learner from the concepts being taught in the instructional multimedia piece. There are three different types of cognitive load: intrinsic, extraneous, and germane (Sweller, Van Merrienboer, & Paas, 1998).

- **Intrinsic cognitive load** - is based on the material being introduced and the experience of the learner.
- **Extraneous cognitive load** – is based on material outside of the content such as presentation methods or activities that force the user to pay attention to multiple sources of information (Sorden, 2005). Examples of extraneous material are: background music, animations flying across the screen, and text on the screen with a narration. Instructional and multimedia designers need to prevent extraneous processing as the more energy the learner spends on trying to process this information the less cognitive capacity they will have for engaging in the learning experience (Mayer & Johnson, 2008).
- **Germane cognitive load** - is based on enhancing the learning experience and results in task resources being devoted to schema acquisition and automation (Sorden, 2005).

Principles of Multimedia

Designers who develop instructional multimedia pieces will need to balance their use of visual and verbal information to effectively engage the learner in the learning process (Bull, 2013). Twelve principles of multimedia design were created by Mayer to help designers balance visual and verbal channels (Mayer, 2009). With the ease of multimedia applications it is more critical than ever for instructional and multimedia designers to understand and apply the twelve principles.

**Twelve principles of multimedia design.** These twelve principles of multimedia design are:

- **Multimedia principle** – Instructional multimedia pieces that contain words and pictures are more effective for learners than just using words. The designer should make sure at least two modes are present: text, video, graphics, animation, and narration (Bull, 2013).
An example could be a screencast of the instructor speaking while showing their PowerPoint presentation on the screen.

- **Spatial contiguity principle** – Multimedia pieces are better designed for the learner when words and pictures are placed near each other vs being a distance apart (Sorden, 2012).
- **Temporal contiguity principle** – Learners respond better to instructional multimedia pieces that present words and graphics continuously rather than one after another (Sorden, 2012).
- **Coherence principle** – Learners are more successful in grasping the concepts in the instructional multimedia piece when irrelevant elements are not included. The multimedia piece should not contain different concepts on the same frame or slide. The designer also shouldn’t include multiple images on the same frame or slide as this can visually overload the learner.
- **Modality principle** – Graphics and narration are more effective in instructional multimedia pieces than text and graphics on a page.
- **Redundancy principle** – Graphics, narration, and printed text should not all be implemented on a slide/frame. Instead only graphics and narration should be presented. The designer should not include interactive animations if a video is being used as it can compete and distract from the learner’s attention.
- **Individual differences principle** (also known as personalization principle) – Formal style conversations should not be used. Instead the instructor should speak in a conversational style (Sorden, 2012).
- **Signaling principle** – Learners are able to recognize and learn information easier when call outs, arrows, and highlighting is used for key aspects.
- **Segmenting principle** – Learners understand the instructional multimedia piece better when the lesson is broken into user-paced chunks rather than all in one multimedia piece (Sorden, 2012).
- **Pre-training principle** – Instructional multimedia is more effective when learners have pre-training on the objectives and key concepts they will learn about.
- **Voice principle** – Learners are more engaged in the learning process when the voice in a multimedia presentation is human vs. a computer generated voice.
- **Image principle** – The instructor’s image on the screen does not generate more meaningful learning than if the image was not present.

**Three challenges.** Cognitive researchers have identified over the past decade three major challenges multimedia designers face when creating instructional multimedia: extraneous content, extraneous details, and complexity (Ibrahim, 2012). Instructional multimedia pieces that contain extraneous content can overload the learners processing capacity. According to Ibrahim (2012) “empirical studies have found that learners performed better on a problem solving transfer test after reviewing a concise lesson, rather than an expanded lesson” (p. 84).

Having too much information (verbal and visual) within the instructional multimedia piece can drain the learner’s ability to focus on the main idea. Adding headers, highlighting key concepts, and using symbols can help the learner focus on the information they need to learn. If an instructional media piece is too detailed or long this can cause a cognitive overload for the learner and is dependent on the learner’s prior knowledge. The more prior knowledge a learner
has about a lesson or topic the less mental effort the learner needs to apply. This results in the learner having more cognitive capacity left for selecting and organizing new information (Höffler & Leutner, 2007). To prevent cognitive overload from happening the lesson should be chunked into segments. If the visual and on screen text is too far away or appears at separate time, it can create a split-attention effect (Aostinho, Tindall-Ford, & Roodenrys, 2013). A multimedia designer needs to understand these three challenges before he or she can effectively develop a sound instructional multimedia piece.

**How Does a Designer Develop Effective Instructional Multimedia Pieces?**

Mayer has conducted numerous research experiments to determine what elements in instructional multimedia result in the most meaningful learning. His research found that contiguous narration and visual graphics in videos are extremely effective for entry level courses, visual learners, and for introducing complex topics (Berk, 2009). Designers need to understand that learners generate more meaningful learning when the text is presented verbally rather than printed on the screen (Kalyuga, Chandler, & Sweller, 2000). An experiment conducted by Park, Flowerday, & Brünken (2015) found that learners learn better with narration instead of the printed word next to a complex visual. Multimedia designers should place graphics and text near each other. In addition graphics that are decorative or do not contribute to the main objectives of the lesson should not be used.

**Redundancy.** Multimedia designers should leave out redundant information. The spoken text should not appear word for word on the screen. Presenting printed words and a narrative can actually decrease learning due to extra memory load (Oud, 2009). If on screen text is used it needs to add to the learning experience. Under certain circumstances, a limited amount of on-screen text can induce germane load rather than extraneous (Adesope & Nesbit, 2012). In a study Mayer conducted about redundancy, he found that adding short redundant text (two to three words) next to the narrated graphics resulted in improvements on retention of the narration but not on transfer (Mayer & Johnson, 2008). This design principle is called the redundancy principle. Redundant information can cause a cognitive overload because the learner's working memory is being used to process unnecessary information (Toh, Munassar, & Yahaya. 2010). Multimedia designers should weed out unnecessary information (printed text, spoken word, and visual graphics).

**Signaling.** Signaling is a technique that can be used and is one of the multimedia principles that can reduce extraneous material. Extraneous material in an instructional multimedia piece can be redundant text, images, diagrams, etc. Multimedia designers can use headers, overview or objective statements, highlight key words or concepts, use call outs for important areas, as well as arrows and animations transitions to draw the learner's attention. The text typeface can be changed by making the text bold, changing the typeface, or the color. The text can also have animation effects added (fade in, fly in, rotation, and flashing). Without the use of signaling learners might experience cognitive load due to having to hold auditory information in working memory until the learner can find the visual reference on the screen (Kalyuga, 2012). In an experiment Mayer conducted to assess the benefits of signaling he found (2008) "in six out of six experiments on airplane and paper-based lessons on lighting and biology, learners who received signaled lessons performed better on transfer tests than students who did not receive signaled lessons" (p. 764). In an experiment conducted by Scheiter signaling resulted in quicker attention
to signaled diagram elements (Scheiter & Eitel, 2015). It is critical that the multimedia designer use signaling sparingly or it can create a cognitive overload.

**Segmentation.** Segmentation focuses on dividing instructional multimedia pieces into smaller topics that are split up across a lesson. An example would be a Photoshop multimedia lesson that covers the tools, filters, effects, and layers. Instead of presenting all of these lessons in one video, a multimedia designer can break this lesson into smaller segments or chunks. Segmentation allows learners to learn one topic before moving on to the next topic. The learner has control over how much of the video they want and can watch different topics out of order.

Multiple studies have proven that segmentation is effective for beginner students, learning material is conceptually complex, and the presentation pace is rapid (Ibrahim, 2012). In an experiment Mayer conducted he compared how learners did in a test of knowledge transfer with one group of students watching a continuous animation showing how an electric motor works and the other group watching the animation broken into segments (Ibrahim, 2012). The results showed the segmented group was more successful. In another study according to Mayer (2008) “in thirteen out of his fourteen experiments he conducted on instructional multimedia lessons on lighting, ocean waves, and brakes learners performed better on their assessment when the lesson was chunked into small pieces rather than the entire lesson being presented in one video” (p. 763).

**Animation vs static image.** Höfler & Leutner conducted experimental research to determine if instructional animations were more effective than static pictures. Their research found there is an advantage to using animations over static pictures, particularly when the animation connects directly to the topic or lesson being addressed. In order for learners to create meaning from an instructional multimedia piece, learners need to create a mental representation of the content displayed in on the slide (de Koning, Tabbers, Rikers, Paas, 2010). According to Jamet, Gavota, & Quaireau (2008) a study they reviewed discovered "color changes and flashing for cueing resulted in positive effectives in retention, transfer, and text-image matching tasks" (p. 143). In an experiment conducted by Scheiter, Schüler, Gerjets, Huk, Hesse (2014) they found that adding animation to verbal explanations helped learners recall immediate information but did not aid in the transfer part of the experiment. There are times when a static picture can be more effective. The picture must be easy to understand, have limited text, and relate directly to the main objective being taught.

**Control.** Research conducted by Mayer has found that when learners have control over the pace of the multimedia the learning is more meaningful and effective (Oud, 2009). Mayer and Chandler conducted several experiments where learners had control of their instructional multimedia piece. Some experiments were very basic where the learner had control of a pause button. In an experiment conducted by Sage, Bonacorsi, Izzo, & Quirk (2015) only a quarter of learners in the experiment used the click-to-pause feature. Though only a quarter of the learners used the click-to-pause feature, the majority of the learners stated they were happy that the click-to-pause feature was there. Another experiment compared learner-control against the system-controlled version. Their results concluded that the learner-control version had better transfer performance when the system-controlled version (Tabbers and de Koeijer, 2010). Allowing the learner to control the content, help, and pace can result in more meaningful learning and prevent cognitive overload. Having a table of contents in the multimedia instructional design allows the learner to quickly locate specific topics of interest quickly without having to scan through the
whole multimedia piece. Chunking the lessons into smaller segments gives the learner the control to focus their attention on aspects of the lesson he or she prefers.

**Interactivity.** Interactivity describes the interactions between student-to-student, teacher-to-student and student-to-content (Evans & Gibbons, 2007). According to Evans & Gibbons (2007), “interaction involves a sequence of three actions: initiation, response, and feedback (p. 1149)”.

1. **Initiation** – presents the learner with a button or control feature for the learner to push/interact with to move on to the response action.
2. **Response** – the learner presses the button or control features that appeared in the last action.
3. **Feedback** – the next slide or information is presented as a result of the learner pushing the button in the last action sequence.

Evans and Gibbons conducted two experiments to determine if using interactivity enhanced the learning process. One experiment was interactive and the other was non-interactive. The results found that learners who viewed the interactive multimedia piece performed better on the problem-solving test and needed less time to complete the memory and problem-solving tests (Evans & Gibbons, 2007). In an experiment conducted by Chen & Catrambone (2014) they found that learners were more motivated and engaged in the learning process when using interactivity. Interactivity can also help learners generate meaningful learning.

**Engagement and feedback.** Not only should instructional multimedia pieces contain signaling, avoid redundancy, chunked into smaller sections, contain interactivity but should also interactive with the learner and allow them to engage with the lesson being presented. Engagement and feedback needs to be meaningful, help the learner apply what he/she learned, and realistic (Oud, 2009). Engagements can contain multiple choice questions, buttons to click on, objects to move around, and links that go to different slides, pages, or websites. If a multimedia designer creates a multiple choice question after the lesson for the learner to answer it is important that the multiple choice question provides meaningful feedback. If the learner answered the question correctly, the correct response should explain why it is correct. If the learner answered the question incorrectly, the correct answer should be given as well as where the learner should go to review the main concepts covered.

**Screencasts.** Jon Udell in 2004 defined a screencast as a way to present digitally recorded playback of a computer screen output that contained spoken narration (Brown, Luterbach, & Sugar, 2009). Screencasts can fall under the category of animations as animations are defined as moving visuals with spoken narration. Examples of screencasts can be an instructor going through a PowerPoint presentation and showing how to use a feature in a software application. There are several benefits for instructional and multimedia designers to consider when creating screencasts. Per Hartsell and Yuen (2006) online video-based instruction “brings courses alive by allowing online learners to use their visual and auditory senses to learn complex concepts and difficult procedures” (p. 31).

In an experiment conducted by Ali, Zamzuri, Samsudin, Hassan, Sidek (2011) they found that learners learned best when the screencast contained narration, were short, simple (did not contain complex animation), and the learner had low prior knowledge. Palaigeorgiou, & Despotakis (2010) conducted an experiment to determine how effective screencasts were and
inquired students on their attitudes towards screencasts. Students’ stated that screencasts increased their application-specific confidence, were more humanized, and were effective in being able to reproduce the procedures the instructor demonstrated. When an instructional or multimedia designer develops screencast multimedia it is important to consider prior knowledge, redundancy, signaling, segmentation, and control as without these there is danger of overloading the limited cognitive capacity of the learner (Brown, Luterbach, & Sugar, 2009).

Conclusion

We live in a video driven world where adult learners are drawn to instructional multimedia pieces. Instructional and multimedia designers must learn about the cognitive theory of multimedia learning and the different principles the designer should consider and implement to design effective instructional multimedia pieces.

Implications of the literature. Effective instructional multimedia pieces should have detailed graphics and narration that contribute to the key aspects of the lesson, include highlighting and use of arrows to draw the learner's eye to the main concepts, the lesson is broken into segments and the learner can control the user controls (play, stop, and pause), engage the learner, and provide an opportunity at the end of the lesson for the learner to apply what he or she has learned.

Limitations with instructional multimedia. One of the biggest weaknesses is that the majority of the experiments that have been conducted were on instructional multimedia pieces that were less than twenty minutes long and were often chunked into smaller pieces. Instructional multimedia pieces require high levels of cognitive processing to analyze and synthesize the visual and narration information and then place this new knowledge in working memory (Ibrahim, 2012). Mayer’s experiments and recommendations are more effective for inexperienced/novice learners than learners who already have knowledge in the subject. Spanjers et al. (2011) conducted experiments on the effect of segmentation on learners with different levels of prior knowledge in a specific field of study and found that learners with prior knowledge learned equally efficiently from non-segmented and segmented instructional multimedia pieces.

In several of Mayer's experiments he states his research is limited in subject area (scientific multimedia and technical how to's) and that further research on different types of instructional multimedia needs to be conducted. Furthermore additional studies and experiments need to be conducted within non-scientific subjects as well learners who are under 18 years of age. Very few studies have been conducted in the effects of cueing in animation and the experiments that have done have mixed results (De Koning, Tabbers, Rikers, & Paas, 2009). Additional studies also need to be conducted to determine the optimal size of textural information (Kalyuga, S., Chandler, P., & Sweller, J. (2004).

Not only are there several theory based limitations but there are several technical limitations such as time constraints, copyright material, technical support, and access to a server that could limit how effective multimedia software and applications could be in creating instructional multimedia pieces (Bull, 2013). If the instructor is developing the course and doesn’t have access to a
multimedia designer this will make it even harder due to concerns above to create instructional multimedia pieces.

**Discussion and recommendations for future research.** To help instructional and multimedia designers understand the cognitive theory of multimedia learning and the different principles a visual table should be created in Excel or another visual tool application that connects different principles to the most effective multimedia software. Within this visual table a listing of what to avoid doing to prevent cognitive load should be shown as well as recommendations on how to improve the instructional multimedia piece – such as chunking a screencast into small topics. There are several recommendations for future research:

- Examine why a small discrepancy between on-screen text and narration is beneficial whereas a large discrepancy is not.
- If the interactivity effect for memory is reproducible
- The effects of attention guidance on transfer tasks
- Establish more specific instructional guidelines for the optimal size of textural information
- Determine best practices using the multimedia principles with learners who have high prior knowledge
- Experiments need to be conducted in more learning/classroom environments than laboratory environments

**References**


